AN EQUAL OPPORTUNITY EMPLO

OFFICIAL BUSINESS
PENALTY FOR PRIVATE USE, $300

02 1M 0004244939 SEP 25 2009
MAILED FROM ZIPCODE 223

NIXIE 3327 1 03 10/27/09

RETURN TO SENDER
NOT DELIVERABLE AS ADDRESSED
UNABLE TO FORWARD
RETURN TO SENDER

RECEIVED
OCT 29 2009
USPTO MAIL CENTER

1 OF 2

PRIORITY MAIL
UNITED STATES POSTAL SERVICE®
www.usps.com
Label 107R, February 2006

OCT 29 2009
IAP93
PATENT & TRADEMARK OFFICE

UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
10/589,956	09/20/2007	Edward H. Cohen	C007-7022US0	2118

56679 7590 09/25/2009

GOSZ AND PARTNERS LLP
ONE STATE STREET
BOSTON, MA 02109

EXAMINER
HADDAD, MAHER M

ART UNIT	PAPER NUMBER
1644	

MAIL DATE	DELIVERY MODE
09/25/2009	PAPER

Please find below and/or attached an Office communication concerning this application or proceeding.

The time period for reply, if any, is set in the attached communication.

Office Action Summary	Application No.	Applicant(s)	
	10/589,956	COHEN ET AL.	
	Examiner	Art Unit	
	Maher M. Haddad	1644	

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --

Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE 3 MONTH(S) OR THIRTY (30) DAYS, WHICHEVER IS LONGER, FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133). Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

Status

1)☒ Responsive to communication(s) filed on *06 August 2009*.

2a)☐ This action is **FINAL**. 2b)☒ This action is non-final.

3)☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

Disposition of Claims

4)☒ Claim(s) *1-33* is/are pending in the application.

4a) Of the above claim(s) *12-33* is/are withdrawn from consideration.

5)☐ Claim(s) _____ is/are allowed.

6)☒ Claim(s) *1-11* is/are rejected.

7)☐ Claim(s) _____ is/are objected to.

8)☐ Claim(s) _____ are subject to restriction and/or election requirement.

Application Papers

9)☐ The specification is objected to by the Examiner.

10)☐ The drawing(s) filed on _____ is/are: a)☐ accepted or b)☐ objected to by the Examiner.

Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).

Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).

11)☐ The oath or declaration is objected to by the Examiner. Note the attached Office Action or form PTO-152.

Priority under 35 U.S.C. § 119

12)☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).

a)☐ All b)☐ Some * c)☐ None of:

1.☐ Certified copies of the priority documents have been received.

2.☐ Certified copies of the priority documents have been received in Application No. _____.

3.☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

* See the attached detailed Office action for a list of the certified copies not received.

Attachment(s)

1) ☒ Notice of References Cited (PTO-892)
2) ☐ Notice of Draftsperson's Patent Drawing Review (PTO-948)
3) ☐ Information Disclosure Statement(s) (PTO/SB/08) Paper No(s)/Mail Date _____.
4) ☐ Interview Summary (PTO-413) Paper No(s)/Mail Date. _____.
5) ☐ Notice of Informal Patent Application
6) ☐ Other: _____.

DETAILED ACTION

1. Claims 1-33 are pending.

2. Applicant's election with traverse of Group I, claims 1-11, drawn to a protein comprising HC and LC that binds to an activated conformation of LFA- 1, filed on 8/06/009, is acknowledged.

Applicant's traversal is on the grounds that the groups of claims are interrelated and can be conveniently searched in one location. In this regard, it is noted that the search class and subclass for each group has not been specifically identified in the restriction requirement. This is not found persuasive because Applicant's inventions do not contribute a special technical feature when viewed over the prior art they do not have a single general inventive concept and so lack unity of invention as set forth in the previous Office Action.

The requirement is still deemed proper and is therefore made FINAL.

3. Claims 12-33 are withdrawn from further consideration pursuant to 37 CFR 1.142(b), as being drawn to nonelected inventions.

4. Claims 1-11 are under examination as they read on a protein comprising HC and LC that binds to an activated conformation of LFA- 1.

5. There appear to be discrepancy between the Sequence Listing and the specification. For example:

A. Claim 5 recites (Xa-S-X2-D-X4-X5-S-X7-A-X8-X9-X10-X11 as SEQ ID NO: 4 (a 13 amino acid (aa) sequence), however, the Sequence Listing lists that SEQ ID NO: 4 is a 31 amino acid sequence

B. The specification on pages 88 and 89, indicates that SEQ ID NO: 41 and 42 is a nucleic acid sequence, however, the Sequence Listing lists SEQ ID NO: 41 and 42 as an amino acid sequence.

C. The specification discloses that SEQ ID NO: 61 (477 aa long) at positions 114-115 are MV (see page 96), however, the Sequence listing list SEQ ID NO: 61 (476 aa long) at the same positions as N.

D. The specification discloses that SEQ ID NO: 60 at position 116 is I (see page 95), however, the Sequence listing list SEQ ID NO: 60 at the same positions as L.

E. The specification discloses that SEQ ID NO: 37 (117 aa long) (see page 86), however, the Sequence listing list SEQ ID NO: 37 (120 aa long) by adding positions 113-115 as XAA.

F. The specification discloses that SEQ ID NO: 36 (see page 86) with positions 59-60 as YY, however, the Sequence listing list SEQ ID NO: 36 with positions 59-60 as TI.

G. The specification discloses that SEQ ID NO: 35 (106 aa long) (see page 85), however, the Sequence listing list SEQ ID NO: 35 (114 aa long) by adding multiple amino acids.

H. The specification discloses that SEQ ID NO: 34 (111 aa long) (see page 85), however, the Sequence listing list SEQ ID NO: 34 (110 aa long) by deleting an amino acid.

I. The Sequence Listing refers to SEQ ID NO: 33 as a 115 amino acid sequence, wherein Xaa can be any naturally occurring amino acid. While in the specification, SEQ ID NO:33 is represented on page 85 as a 109 amino acid sequence with no Xaa.

J. Applicant is required to check all the sequences in the specification and claims to make sure that they do correspond to the sequences in the Sequence Listing.

6. Claim 11 is objected to for the following informalities: the claim number "11," should be "11."

7. The following is a quotation of the second paragraph of 35 U.S.C. 112.

The specification shall conclude with one or more claims particularly pointing out and distinctly claiming the subject matter which the applicant regards as his invention.

8. Claims 1-11 are rejected under 35 U.S.C. 112, second paragraph, as being indefinite for failing to particularly point out and distinctly claim the subject matter which applicant regards as the invention.

A) Claims 1-4 are indefinite in the recitation of "D2-57, DX-2001, C1-54 or P1-G10" because its characteristics are not known. The use of " D2-57, DX-2001, C1-54 or P1-G10" Fab antibody as the sole means of identifying the claimed antibody renders the claim indefinite because " D2-57,DX-2001, C1-54 or P1-G10" is merely a laboratory designation which does not clearly define the claimed product, since different laboratories may use the same laboratory designations to define completely distinct hybridomas or cell lines. It is suggested that a deposit number be cited in the claims.

B) The recitation "D2-57, DX-2001, C1-54 or P1-G10 antibody" in claims 1-4 is ambiguous. Given that D2-57 antibody was fished from Fab phage library (see page 83, line 24), it is not clear whether the D2-57 recited in the claim refers to antibody fragment (Fab) or refers to the converted D2-57 IgG1 recited on page 92, line 8.

C) Claim 1(v-vi) is indefinite, ambiguous and unclear. It is unclear how the nucleic acid that hybridizes (antisense) to a sequence that encodes the heavy/light chain domain of the D2-57, DX-2001, C1-54 or P1-G10 (sense) would encode a heavy/light chain variable domain sequence. The resultant antisense sequence would not encode a heavy/light chain variable domain.

D) The recitation of "hybridizing under stringent conditions" in claim 1(v-vi) is ambiguous. Although the specification discloses on page 30, 2nd ¶ general parameters for calculating such conditions, in the absence of a clear definition of the metes and bounds of this phrase it is unclear which conditions are actually claimed. It is suggested that Applicant amend the claims to recite a particular set of hybridization and wash conditions, such as those exemplified on page 30, 2nd ¶ of the specification, to overcome this rejection.

E) The recitation of percentage identity "80/85/90% identical" in claims 1(iii-iv), 3, 4, without setting a structural feature for the comparison is indefinite.

F) The recitation "Xa-S-X2-D-X4-X5-S-X7-A-X8-X9-X10-X11 (SEQ ID NO: 4)" in claim 5 is indefinite. There are two different SEQ ID NO: 4, the one listed in the claim and the one listed in the Sequence Listing. It is not clear which SEQ ID NO: 4 is being claimed.

9. The following is a quotation of the first paragraph of 35 U.S.C. 112:
The specification shall contain a written description of the invention, and of the manner and process of making and using it, in such full, clear, concise, and exact terms as to enable any person skilled in the art to which it pertains, or with which it is most nearly connected, to make and use the same and shall set forth the best mode contemplated by the inventor of carrying out his invention.

10. Claims 1-11 are rejected under 35 U.S.C. 112, first paragraph, as containing subject matter which was not described in the specification in such a way as to enable one skilled in the art to which it pertains, or with which it is most nearly connected, to make and/or use the invention.

It is apparent that the cell lines that produces the anti-activated LFA-1 Fabs, D2-57, DX-2001, C1-54 are P1-G10 are required to practice the claimed invention. As a required element, it must be known and readily available to the public or obtainable by a repeatable method set forth in the specification. If it is not so obtainable or available, the enablement requirements of 35 USC 112, a deposit of the cell line, which produces this antibody, may satisfy first paragraph. See 37 CFR 1.801-1.809.

If the deposits have been made under the terms of the Budapest Treaty, an affidavit or declaration by applicants or someone associated with the patent owner who is in a position to make such assurances, or a statement by an attorney of record over his or her signature, stating that the cell line has been deposited under the Budapest Treaty and that

the cell lines will be irrevocably and without restriction or condition released to the public upon the issuance of a patent would satisfy the deposit requirement made herein. See 37 CFR 1.808. Further, the record must be clear that the deposit will be maintained in a public depository for a period of 30 years after the date of deposit or 5 years after the last request for a sample *or for the enforceable life of the patent whichever is longer.* See 37 CFR 1.806. If the deposit has not been made under the Budapest treaty, then an affidavit or declaration by applicants or someone associated with the patent owner who is in a position to make such assurances, or a statement by an attorney of record over his or her signature must be made, stating that the deposit has been made at an acceptable depository and that the criteria set forth in 37 CFR 1.801-1.809, have been met.

If the deposits were made after the effective filing date of the application for a patent in the United States, a verified statement is required from a person in a position to corroborate that the cell line described in the specification as filed are the same as that deposited in the depository. Corroboration may take the form of a showing of a chain of custody from applicant to the depository coupled with corroboration that the deposit is identical to the biological material described in the specification and in the applicant's possession at the time the application was filed.

Further, amendment of the specification to disclose the date of deposit and the complete name and address of the depository (ATCC.10801 University Boulevard, Manassas, VA 20110-2209) is required as set forth in 37 C.F.R. 1.809(d).

11. Claims 1-11are rejected under 35 U.S.C. 112, first paragraph, as failing to comply with the written description requirement. The claim(s) contains subject matter which was not described in the specification in such a way as to reasonably convey to one skilled in the relevant art that the inventor(s), at the time the application was filed, had possession of the claimed invention.

Applicant is in possession of an anti-activated LFA-1 antibody comprising VH and/or VL of SEQ ID NOS: 34/33, 36/35, 38/37 or 61/60 (listed on page 85-86 of the specification); or an anti-activated LFA-1 antibody comprising VH CDR1-3 of SEQ ID NOS: 1-3 and VL CDR1-3 of SEQ ID NOs:7-9. or an anti-activated LFA-1 antibody D2-57, DX-2001, C1-54 or P1-G10 (once the deposit is satisfied).

Applicant is not in possession of the protein claimed in claims 1-11.

The scope of the claim encompasses antibodies with 6 intact CDRs as well as a subgenus of antibodies that encompass variation (fragments and/or analogs) in the 6 CDRs and framework. A subgenus of antibodies that encompass up to 10% to 20% variation in the VH and/or VL. A subgenus of antibodies that encompass up to 9 amino acid variations in the VH CDR3 of SEQ ID NO:4. A subgenus of antibodies that encompass less than the full amino acid sequence of the VH CDR1-3 and/or VL CDR1-3 (i.e., 3/5, 13/17, 8/11, 7/11, 4/7 and/or 5/8, respectively.

Prior art discloses 6 CDRs as being essential structure of the antibody's binding site, and thus when intact, would provide enough structure to define the antibody's binding site (structure / function correlation) e.g. where amino acid substitutions can be made so as to change (e.g. 6 CDR's) or retain (e.g. constant or variable framework) antigen binding. However, prior art teaches that variation(s) within the CDRs render antigen binding unpredictable. Therefore, a single antibody species (D2-57) would not be deemed by one of skill in the art to be representative of a claim that defines an antibody that binds activated LFA-1 comprising at least 80-90% identity to the VH and VL chains in D2-57, DX-2001, C1-54 or P1-G10.

The specification provides four anti-activated LFA-1 antibody which was not random combinations of VH and VL i.e., it had specific VH domain (SEQ ID NO: 36-38 and 61) paired with specific VL domain (SEQ ID NO: 33-35 and 60). No other VH domain was provided that share the less than the full length of all CDR1-3 of SEQ ID NO: 1-3 or the full length of all VL CDR1-3 of SEQ ID ON:7-9. The state of the prior art (see e.g. Klimka et al., British Journal of Cancer (2000) 83:252-260, and Beiboer et al., J. Mol., Biol. (2000) 296:833-849) is that methods for screening rely on a two step process where each step results in an antibody. However, each step requires one of the variable domains to be a defined sequence and the defined variable domain provides enough structure to obtain an antibody. The prior art methods do not result in an antibody solely by keeping only one CDR in the VH/VL defined and randomized the rest of the VH and VL domains. The prior art indicated that, in some instances, the CDR3 region is important. However, this region is not solely responsible for binding. The conformation of the other CDRs, as well as framework residues influence binding.

However, neither the specification, nor the prior art provides any examples to support the premise that only one CDR of the VH or VL is solely responsible for antigen binding. The prior art does not support a definition of an antibody structure solely by defining the CDR sequence of a VH or VL. Accordingly, the disclosed species would not be deemed by one of skill in the art to be representative of the claim scope. The claims do not meet the requirement s of 35 USC 112, first paragraph for written description.

The claims encompass antibodies in which modification of the amino acids may vary in either or both the VH and/or VL regions of D2-57, DX-2001, C1-54 or P1-G10 via addition, deletion, substitution or insertion of one or more amino acids.

The instant application encompasses (but does not exemplify) fragments and analogs (deletion/addition/substitution) to the claimed antibody. The specification discloses that the variation in the VL/VH depends on the germline sequence. There is teaching identifying what amino acids can be varied within the VL or VH antibody regions and still retain specific binding member that binds activated LFA-1. However, none of these positions corresponds to claimed antibodies (it is not clear what numbering system Applicant uses). For Example, the specification on page 90, lines 12-16 discloses that an antibody can include a D2-57 light chain with one or more of the following substitutions (or insertion), e.g., at positions:G30S, L40P, A46L, L80P, W96ins, and S97T. However

none of these positions correspond to the D2-57 light chain of SEQ ID NO: 33, depicted on page 85 of the specification. Similarly, the positions of C1-54 and P1-G10, light chain, on page 91, lines 1-2 and 6-7 do not correspond to SEQ ID NOs: 34 and 35. Importantly, none of these variations alone or in combination with other variations have been shown to provide binding to the activated LFA-1.

Moreover, Brown et al (J. Immuno. 1996 May, 3285-91 at 3290 and Tables 1 and 2) describes how one amino acid change in the VH CDR2 of a particular antibody was tolerated whereas, the antibody lost binding upon introduction of two amino changes in the same region. Vajdos et al. (J. Mol. Biol. 2002, Jul 5, 320(2):415-28 at 416) teach that amino acid sequence and conformation of each of the heavy and light chain CDRs are critical in maintaining the antigen binding specificity and affinity which is characteristic of the parent immunoglobulin. Aside from the CDRs, the Fv also contains more highly conserved framework segments which connect the CDRs and are mainly involved in supporting the CDR loop conformations, although in some cases, framework residues also contact antigen. The scope of the claims encompasses antibodies with VH or VL that encompass variation (addition, deletion, substitution) in their FW and CDRs. The prior art discloses that 6 CDRs as being essential structure of antibody's binding site, and thus when intact, would provide enough structure to define the antibody's binding site (structure/function correlation) e.g., where amino acid substitutions can be made so as to change (e.g. 6CDR's) or retain (e.g., constant or variable framework) antigen binding. Neither the prior art not applicant's disclosure defines sufficient representative antibodies and/or sufficient structure/function correlation between modifying the VL or VH regions of the disclosed antibody and the retention of a specific binding member that binds activated LFA-1 to satisfy the WD requirement for the claims.

Claims 5 which is directed to a protein comprising the heavy chain variable domain sequence comprising SEQ ID NO:4 which has 9 variation out of 11 amino acids. The specification on page 92 provides exemplary variants that in the CDR3 region of the heavy chain from residues 96-120 of the DX-2001. The specification discloses that the aspartic acid at position 3 in CDR3 may interact with an Mg2+ ion bound to I-domain. This aspartic acid was conserved 75 of 80 different affinity matured Fabs (see page 93, lines 24-26). However, none of these variants have been shown to provide binding to the activated LFA-1 conformation. These variants had no specific VL domains paired with specific VH domains. Prior art methods do not result in an antibody solely by keeping CDR3 in the VH defined and randomizing the rest of the VH and VL domains.

The Guidelines for the Examination of Patent Applications Under the 35 U.S.C. 112, ¶ 1 "Written Description" Requirement make clear that the written description requirement for a claimed genus may be satisfied through sufficient description of a representative number of species by actual reduction to practice, reduction to drawings, or by disclosure of relevant, identifying characteristics, i.e., structure or other physical and or chemical properties, by functional characteristics coupled with a known or disclosed correlation between function and structure, or by a combination of such identifying characteristics,

sufficient to show the applicant was in possession of the genus (Federal Register, Vol. 66, No. 4, pages 1099-1111, Friday January 5, 2001, see especially page 1106 3rd column).

Vas-Cath Inc. v. Mahurkar, 19 USPQ2d 1111, makes clear that "applicant must convey with reasonable clarity to those skilled in the art that, as of the filing date sought, he or she was in possession of the invention. The invention is, for purposes of the written description inquiry, whatever is now claimed." (See page 1117.) The specification does not "clearly allow persons of ordinary skill in the art to recognize that [he or she] invented what is claimed." (See Vas-Cath at page 1116.). Consequently, Applicant was not in possession of the instant claimed invention. See University of California v. Eli Lilly and Co. 43 USPQ2d 1398.

Applicant is directed to the Guidelines for the Examination of Patent Applications Under the 35 U.S.C. 112, ¶ 1 "Written Description" Requirement, Federal Register, Vol. 66, No. 4, pages 1099-1111, Friday January 5, 2001.

Applicant is invited to point to clear support or specific examples of the claimed invention in the specification as-filed.

12. Claims 1-11 are rejected under 35 U.S.C. 112, first paragraph, because the specification, while being enabling for an anti-activated LFA-1 antibody comprising VH and/or VL of SEQ ID NOS: 34/33, 36/35, 38/37 or 61/60 (listed on page 85-86 of the specification); or an anti-activated LFA-1 antibody comprising VH CDR1-3 of SEQ ID NOS: 1-3 and VL CDR1-3 of SEQ ID NOs:7-9; or an anti-activated LFA-1 antibody produced by the cell line D2-57, DX-2001, C1-54 or P1-G10 (once the deposit is satisfied), does not reasonably provide enablement for the protein claimed in claims 1-11. The specification does not enable any person skilled in the art to which it pertains, or with which it is most nearly connected, to make and use the invention commensurate in scope with these claims.

Factors to be considered in determining whether undue experimentation is required to practice the claimed invention are summarized *In re Wands* (858 F2d 731, 737, 8 USPQ2d 1400, 1404 (Fed. Cir. 1988)). The factors most relevant to this rejection are the scope of the claim, the amount of direction or guidance provided, the lack of sufficient working examples, the unpredictability in the art and the amount of experimentation required to enable one of skill in the art to practice the claimed invention.

The knowledge in the art of making the genus of antibody that binds activated LFA-1 using a set of particular VH or VL CDRs as the starting point is low.

The scope of the claim encompasses antibodies with 6 intact CDRs as well as a subgenus of antibodies that encompass variation (fragments and/or analogs) in the 6 CDRs and framework. A subgenus of antibodies that encompass up to 10% to 20% variation in the VH and/or VL. A subgenus of antibodies that encompass up to 9 amino acid variations in the VH CDR3 of SEQ ID NO:4. A subgenus of antibodies that encompass less than

the full amino acid sequence of the VH CDR1-3 and/or VL CDR1-3 (i.e., 3/5, 13/17, 8/11, 7/11, 4/7 and/or 5/8, respectively.

However, neither the instant specification nor the prior art provide sufficient guidance or direction for one of ordinary skill in the art to make the antibodies encompassed by the breadth of the instant claims.

With respect to making the genus of anti-activated LAF-1 antibodies using a set of particular VH CDRs and/or VL CDRs as the starting point, e.g., SEQ ID NOs: 1-3 and/or CDR 7-9 as recited in claim 1, it is known in the art that antibody-antigen affinity and specificity is a function of not only direct CDR to antigen interactions, but also the interactions of the CDRs with framework residues in the same chain, e.g., VH CDR binding to VH framework residues, and in the opposing chain, e.g., VH CDR binding to VL framework residues. In addition, the CDR residues of each chain can interact with the CDRs of the opposite chain. It is for this reason that antibody humanization protocols, e.g., humanization of a murine antibody, provide extensive guidelines as to the retention of certain murine residues in the context of the human framework so as to preserve this web of interactions, the loss of any one of these interactions having the potential to ablate antibody-antigen binding (see, e.g., Eduardo Padlan, Mol Immunol. 1994 Feb;31(3):169-217, in particular column bridging paragraph on page 177; page bridging paragraph pages 178-179 through page 180; pages 201, 204 and Tables 8, 22 and 23 and Adair et al., United States Patent No. 5,859,205, in particular columns 1-6, 9-11 and 27-28).

It is also known that given one specified variable domain, either heavy or light, the skilled artisans can screen libraries to identify other variable domains that will pair with the starting variable domain and maintain antigen specificity (Portolano et al., J Immunol. 1993 Feb 1;150(3):880-7, see entire document, particularly figure 1). Thus, it is known in the art that artisans can screen for other variable domains that will ensure a functional antibody of defined antigen specificity if a full variable domain is used in the screening assay.

In the instant case, the claims recite only one of the 6 CDRs of a variable domain, not the all the 6 CDRs or the variable domain itself. While CDRs are important for binding and contribute the majority of contact residues with the target antigen, the framework residues are also essential for maintaining the proper antigen-binding conformation of the CDRs and for proper association of the heavy and light chain variable regions.

As such, it appears that making the claimed genus of antibodies would be an unpredictable endeavor requiring far more than routine experimentation because 3 CDRs comprise less than a majority of the residues important for antigen recognition, let alone a single CDR or a CDR variant. Moreover, art techniques for identifying other variable domains by screening require an intact variable domain comprising CDRs interspersed between frameworks as the starting structure to be taken through the screening assay. The instant claims recite less than this minimum structure that is required for screening,

and the instant specification fails to provide sufficient direction or guidance as to the breadth of the frameworks that can accommodate the claimed CDRs while simultaneously providing appropriate structure to pair with a light chain variable domain capable of acting with heavy chain variable domain to create an activated LFA-1 binding site.

The state of the prior art is such that it is well established in the art that the formation of an intact antigen-binding site of antibodies routinely requires the association of the complete heavy and light chain variable regions of a given antibody, each of which consists of three CDRs or hypervariable regions, which provide the majority of the contact residues for the binding of the antibody to its target epitope (Paul, Fundamental Immunology, 3rd Edition, 1993, pp. 292-295, under the heading "Fv Structure and Diversity in Three Dimensions"). The amino acid sequences and conformations of each of the heavy and light chain CDRs are critical in maintaining the antigen binding specificity and affinity, which is characteristic of the immunoglobulin. It is expected that all of the heavy and light chain CDRs in their proper order and in the context of framework sequences which maintain their required conformation, are required in order to produce an antibody having antigen-binding function and that proper association of heavy and light chain variable regions is required in order to form functional antigen binding sites (Paul, page 293, first column, lines 3-8 and line 31 to column 2, line 9 and lines 27-30). Even minor changes in the amino acid sequences of the heavy and light variable regions, particularly in the CDRs, may dramatically affect antigen-binding function as evidenced by Rudikoff et al (Proc. Natl. Acad. Sci. USA, 79(6):1979-1983, March 1982). Rudikoff et al teach that the alteration of a single amino acid in the CDR of a phosphocholine-binding myeloma protein resulted in the loss of antigen-binding function.

While there are some publications, which acknowledge that CDR3 is important, the conformations of other CDRs as well as framework residues influence binding. MacCallum et al (J. Mol. Biol., 262, 732-745, 1996) analyzed many different antibodies for interactions with antigen and state that although CDR3 of the heavy and light chain dominate, a number of residues outside the standard CDR definitions make antigen contacts (see page 733, right col.) and non-contacting residues within the CDRs coincide with residues as important in defining canonical backbone conformations (see page 735, left col.). The fact that not just one CDR is essential for antigen binding or maintaining the conformation of the antigen binding site, is underscored by Casset et al (Biochemical and Biophysical Research Communications, 307:198-205, 2003), which constructed a peptide mimetic of an anti-CD4 monoclonal antibody binding site by rational design and the peptide was designed with 27 residues formed by residues from 5 CDRs (see entire document). Casset et al also states that although CDR H3 is at the center of most if not all antigen interactions, clearly other CDRs play an important role in the recognition process (page 199, left col.) and this is demonstrated in this work by using all CDRs except L2 and additionally using a framework residue located just before the H3 (see page 202, left col.). Thus, the state of the art recognized that it would be highly unpredictable that a specific binding member comprising an antibody variable region but

comprising less than all six CDRs of a parental antibody. Thus, the minimal structure which the skilled artisan would consider predictive of the function of binding the activated LFA-1 includes six CDRs (three from the heavy chain variable region and three from the light chain variable region) from parental antibody D2-57 in the context of framework sequences which maintain their correct spatial orientation have the requisite activated LFA-1 binding function. One of ordinary skill in the art could not predictably extrapolate the teachings in the specification, limited to antibodies that comprise both the heavy chain variable region and the light chain variable region or all six CDRs (i.e., SEQ ID Nos:1-3 and 7-9) of D2-57 that binds the activated LFA-1 to make and use antibodies that comprise fewer than all six CDRs from parental antibody D2-57 (i.e., SEQ ID Nos:33-34), i.e., antibodies comprising a heavy/light chain variable region each having less than the required three CDRs as broadly as claimed. In cases involving unpredictable factors, such as most chemical reactions and physiological activity, more may be required. *In re Fisher*, 427 F.2d 833, 839, 166 USPQ 18, 24 (CCPA 1970) (contrasting mechanical and electrical elements with chemical reactions and physiological activity). See also *In re Wright*, 999 F.2d 1557, 1562, 27 USPQ2d 1510, 1513 (Fed. Cir. 1993); *In re Vaeck*, 947 F.2d 488, 496, 20 USPQ2d 1438, 1445 (Fed. Cir. 1991). This is because it is not obvious from the disclosure of one particular species, what other species will work. See MPEP 2164.03. One of skill in the art would neither expect nor predict the appropriate functioning of the anti- activated LFA-1 antibodies as broadly as is claimed.

Claim 5 recite that the protein binds activated LFA-1 has the heavy chain variable domains sequence comprises Xa-S-X2-D-X4-X5-S-X7-A-X8-X9-X10-X11 of SEQ ID NO: 4. However, the highly diverse VH CDR3 loops are the key determinant of specificity in antigen recognition in antibodies, and may allow the isolation of a new specificity. However, the specification fails to show which of the claimed amino acid or their combination would lead to a binding to activated LFA-1 protein. Given that the claimed antibody recognizes a conformational activated LFA-1 protein, the predictability of which amino acid or their combination that would lead to an antibody that would bind to the claimed activated LFA-1. Changes in the CDR3 sequence of the VH would deviate from the original antigen reactivity and specificity.

At issue is whether the claimed protein that binds activated LFA-1 would function as a pharmaceutical composition (intended use) in claim 11. The exemplification is drawn to the use of D2-57 to bind HA cells (cells expressing an LFA-1 with an I-domain locked in the high affinity conformation) relative to LA cells (cells expressing an LFA-1 with an I-domain locked in the low affinity conformation) (see Fig. 1 in particular). In view of the absence of a specific and detailed description in Applicant's specification of how to effectively use the pharmaceutical composition as claimed, and absence of working examples providing evidence which is reasonably predictive that the claimed pharmaceutical composition are effective for in vivo use, and the lack of predictability in the art at the time the invention was made, an undue amount of experimentation would be required to practice the claimed pharmaceutical composition with a reasonable expectation of success.

Reasonable correlation must exist between the scope of the claims and scope of the enablement set forth. In view on the quantity of experimentation necessary the limited working examples, the nature of the invention, the state of the prior art, the unpredictability of the art and the breadth of the claims, it would take undue trials and errors to practice the claimed invention.

13. The following is a quotation of the appropriate paragraphs of 35 U.S.C. 102 that form the basis for the rejections under this section made in this Office action:

A person shall be entitled to a patent unless --

(b) the invention was patented or described in a printed publication in this or a foreign country or in public use or on sale in this country, more than one year prior to the date of application for patent in the United States.

12. Claims 1, 3, 4, 6-11 are rejected under 35 U.S.C. 102(b) as being anticipated by WO 98/23761 A1 as is evidenced by the specification on page 84, lines 1-7.

The `761 publication teaches and claims a humanized anti-CD11a antibody (LFA-1, CD11a/CD18) which binds specifically to human CD11a I-domain having a heavy chain variable domain sequence comprises a CDR2 that comprises at least 13 amino acids form YIWPSGGNTYYADSVKG (i.e., VISGDGGSTYYADSVKG, published SEQ ID NO: 11) and the light chain variable domain sequence comprises a CDR1 that comprises at least 7 amino acids from RASQSIGSYLN (i.e., RASQSISNYLA, see published SEQ ID NO: 13), A CDR2 that comprises at least 4 amino acids from AASSLQS (i.e., AASSLES, see published SEQ ID NO: 14) and a CDR3 that comprises at least 5 amino acids from QQSYSTPS (i.e., QQYNSLPWT, see published SEQ ID NO:15) (see published claims 1-17, 19 in particular). The humanized anti-CD11a antibodies bind to the I-domain of human CD11a with an affinity of about 1 $X10^{-8}$M or stronger. The antibody has an IC50 (nM) value of no more than about 1 nM for preventing adhesion of Jurkate cells to normal human epidermal keratinocytes expressing ICAM-1 or in a mixed lymphocyte response assay (see published claims 3-4). The `761 publication claims a chimeric construct (F(ab)2(lack Fc) domain) and humanized IgG1 of MHM2 (see claims 11-14). The `761 publication teaches pharmaceutical formulations comprising the antibody with physiologically acceptable carriers, excipients or stabilizers (see page 27, lines 20-21). The reference antibody which is derived from MHM24 antibody, binds to both HA (High affinity, activated LFA-1) and LA cells (un-activated, low affinity LFA-1) (see specification on page 84). The humanized antibody would compete with antibody D2-57, DX-2001, C1-54, or P1-G10 for binding to activated LFA-1 in the absence of evidence to the contrary.

Claims 1(iii-vi) are included because the mouse version of the MHM24 (IgG1) antibody (see Fig. 1 A-B of the `761) and claimed mouse D2-57 (IgG1) (see page 92, of the instant specification) share at least 95% homology. Accordingly, a nucleic acid encoding the MHM24 IgG1 antibody would hybridize to the D2-57 IgG1 antibody claimed, in the absence of evidence to the contrary.
The reference teachings anticipate the claimed invention.

13. Claims 1, 6-11 are rejected under 35 U.S.C. 102(b) as being anticipated by US 2002/0123614 A1 (of record).

The `614 teaches anti-LFA- 1 antibody, or an antigen binding fragment thereof (lacks an Fc domain), which selectively binds to an LFA-1 I-domain in the open conformation (see published claim 30) with high affinity fragment ¶59). The referenced antibody would compete with antibodies D2-57, DX-2001, C1-54, P1-G10 in the absence of evidence to the contrary. The `614 teaches a pharmaceutical composition comprising the antibody and a pharmaceutically acceptable carrier (published claim 28). The `614 publication teach that monoclonal antibodies BL5, F8.8, CBRLFA-1/9, May.035, TS1/22 and TS2/6 (full length IgG antibody) strongly inhibited binding of both wild-type and mutant K287C/K294C "activate or open" (¶202). The `614 teaches chimeric and humanized monoclonal antibodies, comprising both human and non-human portions, which can be made using standard recombinant DNA techniques, can also be used in the methods of the present invention (not immunogenic in humans) (¶86).

The reference teachings anticipate the claimed invention.

14. No claim is allowed.

15. Any inquiry concerning this communication or earlier communications from the examiner should be directed to Maher Haddad whose telephone number is (571) 272-0845. The examiner can normally be reached Monday through Friday from 7:30 am to 4:00 pm. A message may be left on the examiner's voice mail service. If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, Ram Shukla can be reached on (571) 272-0735. The fax number for the organization where this application or proceeding is assigned is 571-273-8300.

Information regarding the status of an application may be obtained from the Patent Application Information Retrieval (PAIR) system. Status information for published applications may be obtained from either Private PAIR or Public PAIR. Status information for unpublished applications is available through Private PAIR only. For more information about the PAIR system, see http://pair-direct.uspto.gov. Should you have questions on access to the Private PAIR system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free).

September 17, 2009

/Maher M. Haddad/
Maher M. Haddad, Ph.D.
Primary Examiner
Technology Center 1600

Notice of References Cited	Application/Control No. 10/589,956	Applicant(s)/Patent Under Reexamination COHEN ET AL.	
	Examiner Maher M. Haddad	Art Unit 1644	Page 1 of 3

U.S. PATENT DOCUMENTS

*		Document Number Country Code-Number-Kind Code	Date MM-YYYY	Name	Classification
*	A	US-5,859,205	01-1999	Adair et al.	530/387.3
	B	US-			
	C	US-			
	D	US-			
	E	US-			
	F	US-			
	G	US-			
	H	US-			
	I	US-			
	J	US-			
	K	US-			
	L	US-			
	M	US-			

FOREIGN PATENT DOCUMENTS

*		Document Number Country Code-Number-Kind Code	Date MM-YYYY	Country	Name	Classification
	N	WO 98/23761	10-1997	PCT	JARDIEU et al.	
	O					
	P					
	Q					
	R					
	S					
	T					

NON-PATENT DOCUMENTS

*		Include as applicable: Author, Title Date, Publisher, Edition or Volume, Pertinent Pages)
	U	Klimka et al.,Human anti-CD30 recombinant antibodies by guided phage antibody selection using cell panning. British Journal of Cancer (2000) 83:252-260,
	V	Beiboer et al.,Guided selection of a pan carcinoma specific antibody reveals similar binding characteristics yet structural divergence between the original murine antibody and its human equivalent. J. Mol., Biol. (2000) 296:833-849)
	W	Brown et al. Tolerance of single, but not multiple, amino acid replacements in antibody VH CDR 2: a means of minimizing B cell wastage from somatic hypermutation? J. Immuno. 1996 May, 3285-91
	X	Vajdos et al. Comprehensive functional maps of the antigen-binding site of an anti-ErbB2 antibody obtained with shotgun scanning mutagenesis. J. Mol. Biol. 2002, Jul 5, 320(2):415-28 at 416

*A copy of this reference is not being furnished with this Office action. (See MPEP § 707.05(a).)
Dates in MM-YYYY format are publication dates. Classifications may be US or foreign.

Notice of References Cited	Application/Control No. 10/589,956	Applicant(s)/Patent Under Reexamination COHEN ET AL.	
	Examiner Maher M. Haddad	Art Unit 1644	Page 2 of 3

U.S. PATENT DOCUMENTS

*		Document Number Country Code-Number-Kind Code	Date MM-YYYY	Name	Classification
	A	US-			
	B	US-			
	C	US-			
	D	US-			
	E	US-			
	F	US-			
	G	US-			
	H	US-			
	I	US-			
	J	US-			
	K	US-			
	L	US-			
	M	US-			

FOREIGN PATENT DOCUMENTS

*		Document Number Country Code-Number-Kind Code	Date MM-YYYY	Country	Name	Classification
	N					
	O					
	P					
	Q					
	R					
	S					
	T					

NON-PATENT DOCUMENTS

*		Include as applicable: Author, Title Date, Publisher, Edition or Volume, Pertinent Pages)
	U	Eduardo Padlan, Anatomy of the antibody molecule. Mol Immunol. 1994 Feb;31(3):169-217
	V	Portolano et al., Lack of promiscuity in autoantigen-specific H and L chain combinations as revealed by human H and L chain "roulette". J Immunol. 1993 Feb 1;150(3):880-7
	W	Paul, Fundamental Immunology, 3rd Edition, 1993, pp. 292-295
	X	Rudikoff et al. Single amino acid substitution altering antigen-binding specificity. Proc. Natl. Acad. Sci. USA, 79(6):1979-1983, March 1982

*A copy of this reference is not being furnished with this Office action. (See MPEP § 707.05(a).)
Dates in MM-YYYY format are publication dates. Classifications may be US or foreign.

Notice of References Cited	Application/Control No. 10/589,956	Applicant(s)/Patent Under Reexamination COHEN ET AL.	
	Examiner Maher M. Haddad	Art Unit 1644	Page 3 of 3

U.S. PATENT DOCUMENTS

*		Document Number Country Code-Number-Kind Code	Date MM-YYYY	Name	Classification
	A	US-			
	B	US-			
	C	US-			
	D	US-			
	E	US-			
	F	US-			
	G	US-			
	H	US-			
	I	US-			
	J	US-			
	K	US-			
	L	US-			
	M	US-			

FOREIGN PATENT DOCUMENTS

*		Document Number Country Code-Number-Kind Code	Date MM-YYYY	Country	Name	Classification
	N					
	O					
	P					
	Q					
	R					
	S					
	T					

NON-PATENT DOCUMENTS

*		Include as applicable: Author, Title Date, Publisher, Edition or Volume, Pertinent Pages)
	U	MacCallum et al. Antibody-antigen interactions: contact analysis and binding site topography. J. Mol. Biol., 262, 732-745, 1996
	V	Casset et al. A peptide mimetic of an anti-CD4 monoclonal antibody by rational design. Biochemical and Biophysical Research Communications, 307:198-205, 2003
	W	
	X	

*A copy of this reference is not being furnished with this Office action. (See MPEP § 707.05(a).)
Dates in MM-YYYY format are publication dates. Classifications may be US or foreign.

doi:10.1016/S0022-2836(02)00264-4 available online at http://www.idealibrary.com on IDEAL® J. Mol. Biol. (2002) 320, 415–428

JMB

Comprehensive Functional Maps of the Antigen-binding Site of an Anti-ErbB2 Antibody Obtained with Shotgun Scanning Mutagenesis

Felix F. Vajdos[1], Camellia W. Adams[1], Timothy N. Breece[2] Leonard G. Presta[3], Abraham M. de Vos[1] and Sachdev S. Sidhu[1]*

[1]*Department of Protein Engineering, Genentech Inc. 1 DNA Way South San Francisco CA 94080, USA*

[2]*Department of Process Sciences, Genentech Inc. 1 DNA Way South San Francisco CA 94080, USA*

[3]*Department of Immunology Genentech Inc., 1 DNA Way South San Francisco, CA 94080 USA*

**Corresponding author*

Shotgun scanning combinatorial mutagenesis was used to study the antigen-binding site of Fab2C4, a humanized monoclonal antibody fragment that binds to the extracellular domain of the human oncogene product ErbB2. Essentially all the residues in the Fab2C4 complementarity determining regions (CDRs) were alanine-scanned using phage-displayed libraries that preferentially allowed side-chains to vary as the wild-type or alanine. A separate homolog-scan was performed using libraries that allowed side-chains to vary only as the wild-type or a similar amino acid residue. Following binding selections to isolate functional clones, DNA sequencing was used to determine the wild-type/mutant ratios at each varied position, and these ratios were used to assess the contributions of each side-chain to antigen binding. The alanine-scan revealed that most of the side-chains that contribute to antigen binding are located in the heavy chain, and the Fab2C4 three-dimensional structure revealed that these residues fall into two groups. The first group consists of solvent-exposed residues which likely make energetically favorable contacts with the antigen and thus comprise the functional-binding epitope. The second group consists of buried residues with side-chains that pack against other CDR residues and apparently act as scaffolding to maintain the functional epitope in a binding-competent conformation. The homolog-scan involved subtle mutations, and as a result, only a subset of the side-chains that were intolerant to alanine substitutions were also intolerant to homologous substitutions. In particular, the 610 Å^2 functional epitope surface revealed by alanine-scanning shrunk to only 369 Å^2 when mapped with homologous substitutions, suggesting that this smaller subset of side-chains may be involved in more precise contacts with the antigen. The results validate shotgun scanning as a rapid and accurate method for determining the functional contributions of individual side-chains involved in protein–protein interactions.

© 2002 Elsevier Science Ltd. All rights reserved

Keywords: phage display; protein engineering; combinatorial mutagenesis; antibody; shotgun scanning

Introduction

Monoclonal antibodies have proven invaluable as reagents in biological chemistry, and more recently, as therapeutic agents.[1] The field of antibody engineering is concerned with technologies that can be used to

Present address: L. G. Presta, DNAX Research Institute of Molecular and Cellular Biology, Inc., 901 California Avenue, Palo Alto, CA 94304, USA.

Abbreviations used: BSA, bovine serum albumin; CC, correlation coefficient; CDR, complementarity determining region; CDR-Hn, (where $n = 1, 2,$ or 3), heavy chain CDR 1, 2, or 3; CDR-Ln, (where $n = 1, 2,$ or 3), light chain CDR 1, 2, or 3; cP3, C-terminal domain of the M13 bacteriophage gene-3 minor coat protein; ECD, extracellular domain; ELISA, enzyme-linked immunosorbant assay; Fab, antigen-binding fragment; Fv, variable fragment; PBS, phosphate-buffered saline; rmsd, root mean square deviation; wt, wild-type.

E-mail address of the corresponding author: sidhu@gene.com

0022-2836/02/$ - see front matter © 2002 Elsevier Science Ltd. All rights reserved

Table 1. Shotgun scanning codons

	Alanine-scan[a]				Homolog-scan[b]	
Wild-type[c]	Codon[d]	m1	m2	m3	Codon[d]	m4
A	GST	G			KCT	S
C	KST	A	G	S	TSC	S
D	GMT	A			GAM	E
E	GMA	A			GAM	D
F	KYT	A	S	V	TWC	Y
G*	GST	A			GST	A
H	SMT	A	D	P	MAC	N
I*	RYT	A	T	V	RTT	V
K	RMA	A	E	T	ARG	R
L	SYT	A	P	V	MTC	I
M	RYG	A	T	V	MTG	L
N*	RMC	A	D	T	RAC	D
P*	SCA	A			SCA	A
Q*	SMA	A	E	P	SAA	E
R	SST	A	G	P	ARG	K
S*	KCC	A			KCC	A
T	RCT	A			ASC	S
V	GYT	A			RTT	I
W	KSG	A	G	S	TKG	L
Y	KMT	A	D	S	TWC	F

For each scan, degenerate shotgun codons were designed to encode the wild-type amino acid and one or more substitutions. Asterisks (*) indicate wild-type amino acid residues for which both the alanine and homolog-scan codons encode a common substitution.

[a] The shotgun alanine-scan codon for each amino acid ideally encodes only the wild-type or alanine (m1), but the nature of the genetic code necessitates the occurrence of two other amino acid residues (m2 and m3) for some substitutions. In the case of wild-type alanine, the shotgun codon was designed to encode alanine and glycine.

[b] For the homolog-scan, binomial shotgun codons were designed to encode the wild-type and a similar amino acid (m4).

[c] Amino acid residues are represented by the single letter amino acid code.

[d] Equimolar DNA degeneracies in shotgun codons are represented by the IUB code (K = G/T, M = A/C, R = A/G, S = G/C, W = A/T, Y = C/T).

dissect and rationalize the requirements for antibody structure and function.[2] This knowledge can then be used to improve or alter particular antibody–antigen interactions, or even to engineer completely novel-binding specificities.

The specificity and affinity of an antibody for its cognate antigen is determined by the sequence and structure of the variable fragment (Fv): a heterodimer consisting of the N-terminal domains of the heavy and light chains. Even within the Fv, antigen binding is primarily mediated by the complementarity determining regions (CDRs), six hypervariable loops (three each in the heavy and light chains) which together present a large contiguous surface for potential antigen binding. Aside from the CDRs, the Fv also contains more highly conserved framework segments which connect the CDRs and are mainly involved in supporting the CDR loop conformations,[3,4] although in some cases, framework residues also contact antigen.[5,6] As an important step to understanding how a particular antibody functions, it would be very useful to assess the contributions of each CDR side-chain to antigen binding, and in so doing, to produce a functional map of the antigen-binding site.

Site-directed mutagenesis is a powerful tool for mapping binding energetics at protein–protein interfaces.[7,8] In this process, individual DNA codons are systematically altered and the corresponding mutant proteins are expressed, purified, and assayed for activity relative to the wild-type. The effects of individual side-chain substitutions can then be assessed in terms of $\Delta\Delta G_{mut-wt}$, the difference in binding free energy between the mutant and wild-type protein. By analyzing panels of point mutants, a detailed map of the binding energetics can be obtained, but the process can be very laborious because individual mutant proteins must be made and analyzed separately. In particular, a comprehensive analysis of an antigen-binding site would ideally encompass all CDR residues, and this would require the analysis of dozens or even hundreds of point mutants.[9,10]

Recently, a general and rapid combinatorial mutagenesis strategy has been developed for exploring protein structure and function.[11] "Shotgun scanning" mutagenesis uses phage-displayed libraries of protein mutants constructed using degenerate codons with restricted diversity. For example, codons may be chosen to preferentially allow the wild-type (wt) or alanine in the case of a shotgun alanine-scan. The library pool is then subjected to binding selections to enrich for clones that retain affinity for a binding partner, and following selection, DNA sequencing is used to determine the ratio of wild-type/mutant (wt/mut) at each varied position. This ratio can be used to assess binding contributions of each side-chain with good correlation to those obtained with traditional site-directed mutagenesis. The method is very rapid because many side-chains are simultaneously scanned with a single library, and the analysis is based on DNA sequencing which circumvents the need for protein purification and biophysical analysis.

We used the shotgun scanning approach to study the antigen-binding site of a humanized monoclonal antibody (humAb2C4) that binds to the extracellular domain of the human receptor tyrosine kinase ErbB2 (ErbB2-ECD, K_d = 8.5 nM), and in so doing, inhibits tumor growth (C.W.A., unpublished results). The antigen-binding portion of humAb2C4 was displayed on M13 bacteriophage in an Fab format (Fab2C4), i.e. a heterodimer consisting of the light chain and the variable and first constant domains of the heavy chain. We conducted two different shotgun scans, with each scan covering essentially the complete sequences of all six CDRs. With a shotgun alanine-scan, we assessed the effects of removing all side-chain atoms past the β-carbon, fairly drastic mutations that can be used to infer the roles of individual side-chains in protein structure and function.[7] We also conducted a more subtle scan, termed a shotgun homolog-scan, in which we substituted

	24	25	26	27	28	29	30	31	32	33	34							
L1	K	A	S	Q	D	V	S	I	G	V	A							
	50	51	52	53	54	55	56											
L2	S	A	S	Y	R	**Y**	T											
L3	89	90	91	92	93	94	95	96	97									
	Q	Q	**Y**	Y	I	Y	**P**	Y	T									
H1	26	27	28	29	30	31	32	33	34	35								
	G*	F*	T	F*	T	**D**	**Y**	**T**	M	**D**								
H2	50	51	52	52a	53	54	55	56	57	58	59	60	61	62	63	64	65	
	D	V	**N**	**P**	**N**	**S**	**G**	G	S	**I**	Y	N	Q	R	F	**K**	G	
H3	95	96	97	98	99	99a	99b	100	101	102								
	N	**L**	**G**	**P**	S	**F**	**Y**	**F**	**D**	Y								

Figure 1. Sequences of the Fab2C4 CDRs. The sequence of each CDR is shown along with the position of each residue in the numbering scheme† of Kabat *et al.*[37] Residues shown to be important for ErbB2-ECD binding in either the shotgun alanine or homolog-scan are shown in bold or underlined, respectively ($F_{wt/mut} > 10$, see Tables 3 and 4). Asterisks (*) indicate residues that were not analyzed in the shotgun scans.

Table 2. Fab2C4 shotgun scanning libraries

	Mutated regions				Diversity	
Library	CDRs	Residues	Shotgun codons	Mutagenic oligonucleotides	Theoretical	Actual
HAa	H1, H2, H3	T28, T30, D31, Y32, T33, D50, V51, N52, N53, S54, I58, N60, Q61, N95, L96, P98, S99	Alanine	H1-A1, H2-A1, H3-A1	3.3×10^7	1.5×10^{10}
HAb	H1, H2, H3	D35, P52a, G55, G56, S57, Y59, R62, F63, K64, G65, G97, F99a, Y99b, F100, D101, Y102	Alanine	H1-A2, H2-A2, H3-A2	1.7×10^7	2.4×10^{10}
LAa	L1, L2, L3	Q27, D28, S30, I31, G32, S50, S52, Y53, Y55, Y91, Y92, I93, Y94, Y96	Alanine	L1-A1, L2-A1, L3-A1	8.3×10^7	1.4×10^{10}
LAb	L1, L2, L3	K24, A25, S26, V29, V33, A34, A51, R54, T56, Q89, Q90, P95, T97	Alanine	L1-A2, L2-A2, L3-A2	1.6×10^4	2.5×10^{10}
HHa	H1, H3	T28, T30, D31, Y32, T33, M34, D35, N95, L96, G97, P98, S99, F99a, Y99b, F100, D101, Y102	Homolog	H1-H, H3-H	1.3×10^5	2.4×10^{10}
HHb	H2	D50, V51, N52, P52a, N53, S54, G55, G56, S57, I58, Y59, N60, Q61, R62, F63, K64, G65	Homolog	H2-H	1.3×10^5	2.2×10^{10}
LH	L1, L2, L3	K24, A25, S26, Q27, D28, V29, S30, I31, G32, V33, A34, S50, A51, S52, Y53, R54, Y55, T56, Q89, Q90, Y91, Y92, I93, Y94, P95, Y96, T97	Homolog	L1-H, L2-H, L3-H	1.3×10^8	2.4×10^{10}

Libraries were designed to replace the codons for the indicated residues with either alanine-scan or homolog-scan shotgun codons (Table 1). Libraries were constructed using the indicated mutagenic oligonucleotides (see Materials and Methods), and in each case, the theoretical diversity (the number of amino acid combinations encoded by the mutagenic oligonucleotides) was exceeded at least 100-fold by the actual diversity of the constructed library.

each wild-type residue with a similar amino acid, to gain insight into which positions require precise side-chain geometries and chemistry. When the mutagenesis results were mapped onto the three-dimensional crystal structure of Fab2C4, each scan provided a comprehensive view of how the CDR side-chains contribute to the formation of a functional antigen-binding site. The two views are distinct yet complementary: together, they provide a clearer understanding of antibody structure and function than would be possible with either scan alone.

Results

Shotgun alanine-scan of Fab2C4

For the shotgun alanine-scan, we replaced wt codons with degenerate codons that ideally encoded the wt amino acid or alanine (m1 in Table 1), although the nature of the genetic code

† Antibody residues are designated by a letter in lower case italics denoting the heavy or light chain (*h* or *l*, respectively), followed by the amino acid in the one-letter code, followed by the position in the chain. For example, *h*D101 denotes an aspartic acid residue at position 101 in the heavy chain.

Table 3. Fab2C4 light chain shotgun scan

	Wt/mut ratios											
	Antigen selection				Display selection				$F_{wt/mut}$			
Residue[a]	Wt/m1	Wt/m2	Wt/m3	Wt/m4	Wt/m1	Wt/m2	Wt/m3	Wt/m4	m1	m2	m3	m4
K24	0.89	4.2	0.96	0.88	0.42	0.79	0.52	1.0	2.1	5.3	1.8	0.86
A25	3.7			2.8	2.0			1.6	1.8			1.8
S26*	3.5			2.8	2.9			1.5	1.2*			1.9*
Q27*	0.67	1.5	2.5	0.51	0.88	1.2	0.94	0.73	0.76	1.3*	2.7	0.70*
D28	1.1			1.8	0.99			1.9	1.1			1.0
V29	6.1			3.5	2.5			2.0	2.4			1.8
S30*	1.8			1.1	1.5			0.87	1.1*			1.3*
I31*	0.91	2.8	0.57	0.64	1.7	2.7	0.56	0.55	0.53	3.7	1.0*	1.2*
G32*	3.3			4.8	2.9			3.9	1.1*			1.2*
V33	16			3.1	3.3			2.8	4.8			1.1
A34	16			5.5	3.6			2.5	4.6			2.2
S50*	1.0			0.78	1.3			0.87	0.77*			0.89*
A51	1.7			1.6	0.90			0.85	1.9			1.8
S52*	1.3			1.2	1.5			1.7	0.85*			0.70*
Y53	1.9	97	4.4	1.4	1.6	3.5	1.2	1.3	1.2	**28**	3.7	1.1
R54	3.2	4.1	1.9	3.0	1.7	3.7	1.0	2.4	1.8	1.1	1.9	1.3
Y55	32	80	53	4.8	1.4	2.3	0.89	0.95	**23**	**35**	**60**	5.1
T56	0.49			0.88	0.89			0.76	0.55			1.2
Q89*	8.8	10	70	3.6	0.77	2.4	3.4	1.9	11	4.2*	21	1.8*
Q90*	2.4	1.1	>36	0.67	0.88	1.9	2.3	0.71	2.7	0.58*	>16	0.94*
Y91	>166	>166	166	0.94	1.8	3.5	0.97	1.2	>92	>47	138	0.76*
Y92	1.2	3.7	1.1	0.88	1.3	2.1	0.84	0.6	0.96	1.8	0.76	1.5
I93*	1.7	1.6	0.81	0.69	1.7	1.5	0.64	0.53	1.0	1.1	1.3*	1.3*
Y94	6.7	30	5.5	1.3	1.9	3.0	1.7	0.63	3.6	**10**	3.2	2.0
P95*	13			9.7	1.1			1.74	**12***			5.6*
Y96	0.99	>66	2.1	0.36	2.1	18	2.2	0.91	0.48	>3.7	0.95	0.40
T97	0.56			0.28	0.89			0.35	0.62			0.80

For each of the listed light chain residues, the effect of each mutation (Table 1) was assessed using data from either the alanine-scan libraries (m1, m2, and m3) or the homolog-scan libraries (m4) described in Table 2. The wt/mut ratios were determined from the sequences of binding clones isolated after selection for binding to either the ErbB2-ECD (antigen selection) or an anti-tag antibody (display selection). The function ratio ($F_{wt/mut}$) for each mutation was derived by dividing the antigen selection wt/mut ratio by the display selection wt/mut ratio. $F_{wt/mut}$ provides a quantitative estimate of the effect of each mutation on the binding affinity of Fab2C4 for ErbB2-ECD. Deleterious effects are indicated by $F_{wt/mut}$ values greater than 1.0, and mutations that have large deleterious effects ($F_{wt/mut} > 10$) are shown in bold text. In cases where a particular mutation was not observed amongst the antigen selection sequences, only a lower limit could be defined for the wt/mut ratio and the $F_{wt/mut}$ (indicated by a greater than sign). Asterisks (*) indicate residues for which the alanine and homolog-scan codons encoded a common substitution.

[a] Residues are denoted by the single letter amino acid code and are numbered according to the scheme of Kabat *et al.*[37]

necessitated two other amino acid substitutions for some residues (m2 and m3 in Table 1). In positions where alanine was the wt, we used a degenerate codon that encoded alanine or glycine. The six CDRs of Fab2C4 encompass a total of 64 residues (Figure 1). We constructed two libraries (HAa and HAb) that together covered 33 of the 37 heavy chain CDR residues and two libraries (LAa and LAb) that together covered all 27 light chain CDR residues (Figure 1 and Table 2). Each library contained $>10^{10}$ unique members, and thus in each case, the theoretical diversity for combinatorial mutagenesis at the scanned positions was exceeded by at least 100-fold (Table 2).

Phage pools from each library were subjected to two different selections. The first selection (display selection) isolated variants capable of binding to a monoclonal antibody specific for the epitope tag fused to the N terminus of the Fab2C4 light chain. The second selection (antigen selection) isolated variants capable of binding to ErbB2-ECD. Close to 100 binding clones were sequenced from each selection; the sequences were aligned, and at each mutated position, the occurrences of wt or each designed substitution were tabulated (see Materials and Methods for details). For each selection, these data were used to calculate the wt/mut ratio for each mutation at each position (Tables 3 and 4).

Because the wt/mut ratio is the statistical preference for the wt relative to the mutant, it correlates with the effect of each mutation on the selected trait (i.e. binding to the anti-tag antibody or ErbB2-ECD). Ratios greater than or less than 1 indicate deleterious or beneficial mutations, respectively.

The anti-tag antibody selected for phage variants that displayed assembled Fab2C4 fragments containing both the heavy and light chains. This is because the heavy chain was fused directly to a bacteriophage coat protein while the epitope tag was fused to the light chain N terminus. Thus, the anti-tag antibody only binds to phage particles that contain a light chain associated with the

Table 4. Fab2C4 heavy chain shotgun scan

	Wt/mut ratios								$F_{wt/mut}$			
	Antigen selection				Display selection							
Residue[a]	Wt/m1	Wt/m2	Wt/m3	Wt/m4	Wt/m1	Wt/m2	Wt/m3	Wt/m4	m1	m2	m3	m4
T28	4.5			0.94	0.7			0.47	6.4			2.0
T30	0.33			0.27	0.7			0.39	0.47			0.69
D31	170			29	1.4			1.1	**120**			**26**
Y32	>161	>161	>161	17	2.0	3.1	1.1	0.85	**>81**	**>52**	**>150**	**20**
T33	20			8.9	0.94			0.38	**21**			**23**
M34	ND[b]	ND	ND	2.2	ND	ND	ND	0.88	ND	ND	ND	2.5
D35	2.8			14	0.14			0.90	**20**			**15**
D50	170			>91	0.24			0.41	**710**			**>220**
V51	10			1.3	1.1			1.8	9.4			0.73
N52*	>168	168	84	>91	0.41	0.34	0.80	0.83	**>410**	**490***	**110**	**>110***
P52a*	72			14	6.1			0.62	**12***			**23***
N53*	>166	166	>166	>91	1.4	0.97	2.6	0.57	**>120**	**170***	**>64**	**>160***
S54*	84			>91	0.33			1.1	**260***			**>83***
G55*	14			90	0.40			2.9	**34***			**31***
G56*	0.60			0.36	5.0			2.6	0.12*			0.14*
S57*	7.0			0.47	4.4			0.86	1.6*			0.55*
I58*	45	45	4.5	2.1	0.86	0.95	0.51	0.61	**53**	**47**	8.8*	3.4*
Y59	33	>59	9.8	0.78	8.7	10.4	1.8	0.58	3.8	>5.7	5.4	1.3
N60*	4.8	4.4	120	3.0	1.2	0.91	15	1.8	4.0	4.8*	8.0	1.7*
Q61*	2.6	0.98	1.1	0.69	0.53	0.42	2.0	0.71	4.8	2.3*	0.55	0.97*
R62	4.3	>44	4.0	1.3	1.2	15	0.24	1.2	3.6	2.9	17	1.0
F63	26	26	4.6	3.2	6.6	2.2	8.8	4.0	4.4	**12**	0.52	0.81
K64	54	54	6.0	0.57	4.9	7.7	2.7	0.67	**12**	7.0	2.2	0.85
G65*	5.8			9.1	2.50			3.9	2.3*			2.4*
N95*	>170	21	>170	21	1.8	2.0	2.1	3.1	**>98**	**11***	**84**	6.9*
L96	23	>45	0.35	1.5	0.11	0.33	0.19	1.2	**210**	**>140**	1.8	1.3
G97*	>78			89	3.3			2.1	**>24***			**42***
P98*	>178			29	1.9			0.44	**>94***			**65***
S99*	2.8			7.0	0.55			1.6	5.0*			4.4*
F99a	>75	>75	>75	10	2.4	5.4	1.3	1.1	**>31**	**14**	**58**	9.1
Y99b	>74	74	74	1.7	0.8	4.1	1.7	0.49	**>93**	**18**	**44**	3.5
F100	77	>77	77	17	2.6	5.9	1.5	5.1	**30**	**13**	**51**	3.3
D101	9.1			>87	1.1			2.5	8.3			**>35**
Y102	8.3	7.5	3.2	2.8	2.3	1.9	2.1	0.92	3.6	3.9	1.5	3.0

For each of the listed heavy chain residues, the effect of each mutation (Table 1) was assessed using data from either the alanine-scan libraries (m1, m2, and m3) or the homolog-scan libraries (m4) described in Table 2. The wt/mut ratios were determined from the sequences of binding clones isolated after selection for binding to either the ErbB2-ECD (antigen selection) or an anti-tag antibody (display selection). The function ratio ($F_{wt/mut}$) for each mutation was derived by dividing the antigen selection wt/mut ratio by the display selection wt/mut ratio. $F_{wt/mut}$ provides a quantitative estimate of the effect of each mutation on the binding affinity of Fab2C4 for ErbB2-ECD. Deleterious effects are indicated by $F_{wt/mut}$ values greater than 1.0, and mutations that have large deleterious effects ($F_{wt/mut} > 10$) are shown in bold text. In cases where a particular mutation was not observed amongst the antigen selection sequences, only a lower limit could be defined for the wt/mut ratio and the $F_{wt/mut}$ (indicated by a greater than sign). Asterisks (*) indicate residues for which the alanine and homolog-scan codons encoded a common substitution.

[a] Residues are denoted by the single letter amino acid code and are numbered according to the scheme of Kabat *et al.*[37]

[b] ND indicates that these values were not determined, because we forgot to include this residue in the alanine-scan libraries.

phage-displayed heavy chain. Most of the wt/mut ratios for the display selection were close to 1.0, indicating that the mutations did not significantly affect Fab2C4 display levels (Tables 3 and 4). However, several mutations exhibited wt/mut ratios significantly greater than 1.0 (e.g. *h*P52aA, *h*Y59A, *h*F63A), suggesting that these mutations reduced display. Conversely, for a few mutations, wt/mut ratios significantly less than 1.0 suggest that these mutations may actually increase display (e.g. *h*D35A, *h*L96A).

In the selection for binding to ErbB2-ECD, mutations could effect the selection either by altering the level of Fab2C4 display (as in the display selection), or alternatively, by directly or indirectly altering the side-chains that make binding contacts with the antigen. In this selection, alanine substitutions at three light chain positions (Table 3) and 21 heavy chain positions (Table 4) exhibited wt/mut ratios greater than 10.

To obtain a quantitative estimate of each mutation's effect on ErbB2-ECD binding affinity, we divided the wt/mut ratio from the antigen selection by the wt/mut ratio from the display selection. This operation corrected for effects on Fab2C4 display and provided a number which we termed the function ratio ($F_{wt/mut}$). As we have shown previously, the $F_{wt/mut}$ value for each mutation is approximately equal to the corresponding ratio of equilibrium binding constants ($K_{a,wt}/K_{a,mut}$),[11] and thus, it provides a good estimate of the effect of each mutation on the

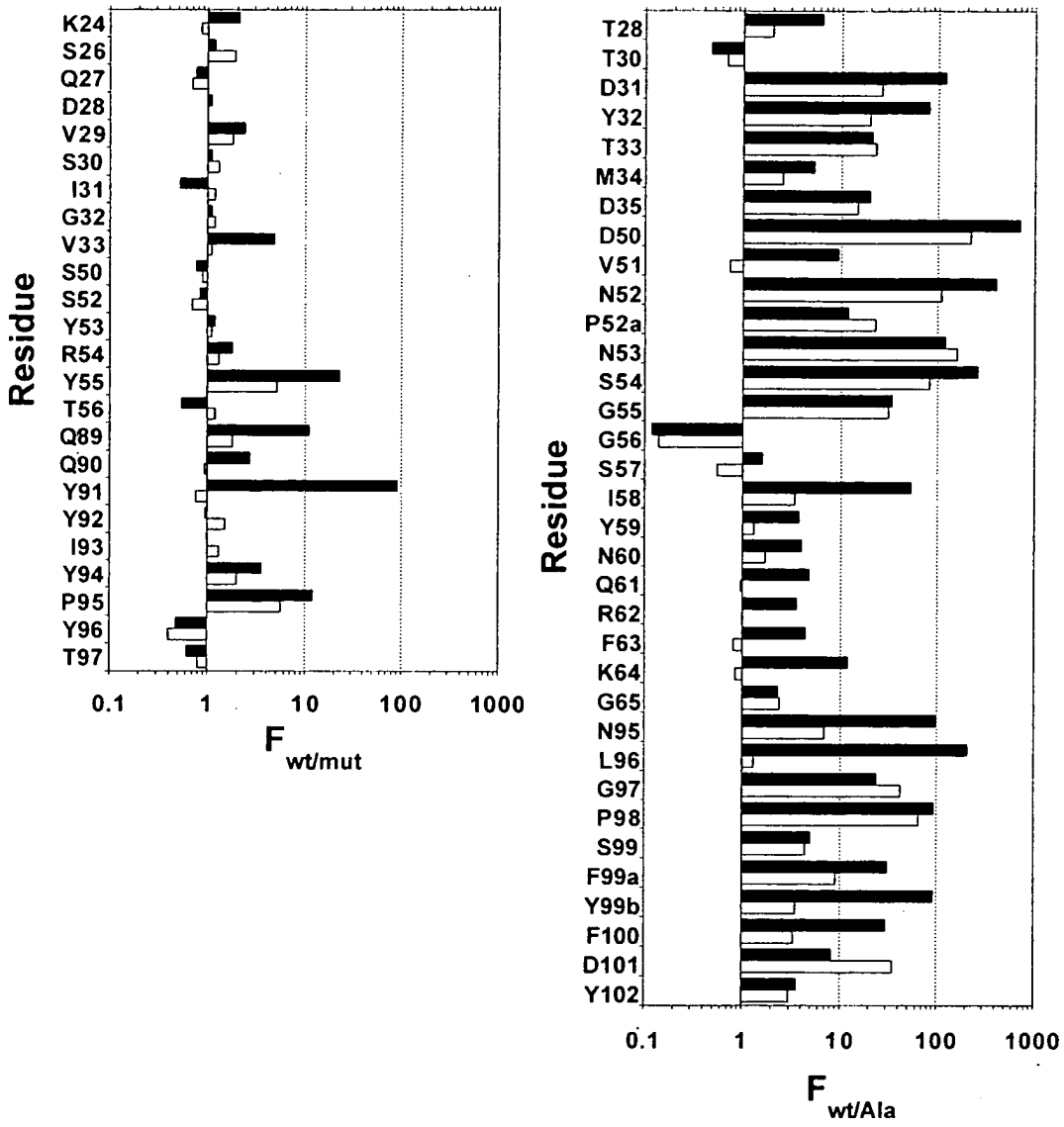


Figure 2. $F_{wt/mut}$ values measuring the effects of Fab2C4 CDR mutations on the binding affinity for ErbB2-ECD. Values are shown for either alanine (black bars) or homolog (white bars) substitutions. Data for (a) the light chain were from Table 3, and data for (b) the heavy chain were from Table 4 (except the mutation *h*M34A for which the $EC_{50,mut}/EC_{50,wt}$-value from Table 5 was plotted).

equilibrium binding constant between Fab2C4 and ErbB2-ECD. Alanine substitutions at three light chain positions and 19 heavy chain positions exhibited $F_{wt/Ala}$ values greater than 10, indicating that side-chains at these positions contribute significantly to the binding affinity of Fab2C4 for ErbB2-ECD (Tables 3 and 4, Figure 2).

Shotgun homolog-scan of Fab2C4

In the shotgun homolog-scan libraries, each scanned position was represented by a binomial codon that encoded only the wild-type and a similar amino acid (Table 1). We constructed two libraries (HHa and HHb) that together covered 34 heavy chain CDR residues and a single library (LH) that covered all 27 light chain CDR residues (Figure 1 and Table 2). As with the alanine-scans, the library diversities were sufficient to exceed the theoretical diversities by at least 100-fold (Table 2).

Each library was subjected to separate selections for binding to anti-tag antibody or ErbB2-ECD and $F_{wt/mut}$ values were determined for each mutation, as described above for shotgun alanine-scanning. The $F_{wt/mut}$ values for many homolog substitutions were significantly lower than those for the corresponding alanine substitutions; no light chain residues and only 13 heavy chain

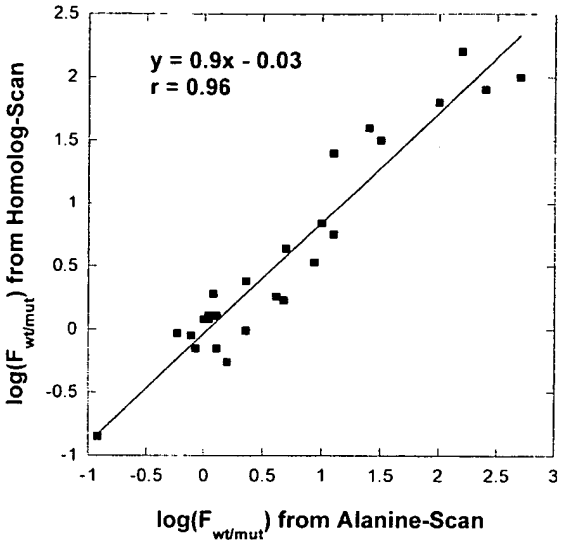


Figure 3. Correlation between $F_{wt/mut}$ values determined using data from the shotgun alanine (x-axis) or homolog-scan (y-axis). The alanine and homolog-scan data could be used to determine the $F_{wt/mut}$ values for 26 identical point mutations that overlapped in the two scans (asterisks in Tables 3 and 4). The logarithms of the $F_{wt/mut}$ values are plotted and the least squares linear fit of the data is shown, with the corresponding equation and Pearson's coefficient (r) given at the top.

residues exhibited $F_{wt/mut}$ values greater than 10 (Tables 3 and 4, Figure 2).

Comparison of $F_{wt/mut}$ values for identical mutations in different libraries

While most substitutions in the homolog-scan were designed to be different from those in the alanine-scan, there was some overlap (residues with asterisks in Table 1). Glycine, proline and serine were substituted with alanine in both scans. Furthermore, for asparagine, glutamine, and isoleucine, the alanine-scan used tetranomial codons that encoded the homolog-scan substitution in addition to alanine and wt. Thus, for 26 mutations that overlapped in the two scans, we could compare the $F_{wt/mut}$ values determined from the alanine-scan to those determined from the homolog-scan (asterisks in Tables 3 and 4). For these identical mutations, a least squares linear fit of the logarithms of the $F_{wt/mut}$ values from the alanine-scan *versus* the logarithms of the $F_{wt/mut}$ values from the homolog-scan showed a strong correlation ($r = 0.96$), with a slope close to 1.0 and a y-intercept close to zero (Figure 3). Thus, it appears that identical point mutations in different combinatorial libraries have very similar effects on the binding affinity of Fab2C4 for ErbB2-ECD.

Table 5. Relative binding activities for Fab2C4 point mutants

Mutant	$EC_{50,mut}/EC_{50,wt}$
*h*M34A	5.3
*h*M34L	1.8
*h*N52A[a]	$>10^3$
*h*N52Q[a]	$>10^3$
*h*N53A[a]	$>10^3$
*h*N53D[a]	$>10^3$
*h*N53Q[a]	$>10^3$
*h*S54A[a]	$>10^3$

The binding activities of mutant proteins were assessed as EC_{50} values, and the ratio of $EC_{50,mut}/EC_{50,wt}$ was determined as a measure of the fold reduction in ErbB2-ECD binding activity due to each point mutation (see Materials and Methods).

[a] For extremely deleterious mutations, EC_{50} values could not be determined because binding could not be saturated. Thus, only a lower limit ($>10^3$) for fold reduction in ErbB2-ECD binding could be estimated for these mutations.

Binding activity measurements with Fab2C4 point mutants

Site-directed mutagenesis was used to construct genes encoding Fab2C4 point mutants; the mutants were expressed in *Escherichia coli* and the recombinant proteins were purified. An enzyme-linked immunosorbant assay (ELISA) with immobilized ErbB2-ECD was used to measure the binding activity of wt Fab2C4 and each mutant protein. For each protein, the EC_{50} was determined as the Fab concentration corresponding to the half-maximal binding signal. By dividing the EC_{50} for each Fab2C4 point mutant by the EC_{50} for wt Fab2C4, we obtained a measure of the fold reduction in ErbB2-ECD binding activity due to each point mutation (Table 5), and these data were in good agreement with the shotgun scanning results (Table 4 and Figure 2(b)). Both methods indicated that mutations at positions *h*N52, *h*N53, and *h*S54 greatly reduced binding affinity for ErbB2-ECD, while the mutation *h*M34L caused only a modest 2-fold reduction.

Three-dimensional structure of Fab2C4

The X-ray crystal structure of Fab2C4 was determined by the molecular replacement method, using as a search model the coordinates of humanized Fab4D5 version 4 (Fab4D5v4),[12] an antibody fragment that also binds to ErbB2-ECD but recognizes an epitope distinct from that recognized by Fab2C4.[13] The structure was refined at 1.8 Å resolution to R_{work} and R_{free} values of 19.7% and 23.0%, respectively. The details of the structure determination and refinement are described in Materials and Methods; data collection and refinement statistics are shown in Table 6. Fab2C4 and Fab4D5v4 share 91% sequence identity; most of the differences reside in the CDRs, as the framework regions differ at only seven positions. Thus, it is not surprising that the Cα atoms of the two structures superimpose with a root mean square deviation (rmsd) of 1.5 Å, excluding the CDRs. Some of this difference can be attributed to "hinge"-motion between the variable and constant

Table 6. Data collection and refinement statistics for Fab2C4

A. *Unit cell*	
Space group	$P2_1$
a (Å)	41.97
b (Å)	64.25
c (Å)	79.44
β (deg.)	105.44
B. *Diffraction data*	
Resolution (Å)	15–1.8 (1.9–1.8)[a]
No. of reflections	85,734
No. of unique reflections	36,884
R_{merge}[b]	0.065 (0.328)[a]
Completeness (%)	97.6 (97.6)[a]
$I/\sigma(I)$	5.2 (1.4)[a]
Redundancy	2.3 (2.3)[a]
C. *Refinement*	
R_{work}[c]	0.197
R_{free}[c]	0.230
No. of protein atoms	3323
No. of water molecules	382
No. of sulfate ions	2
Average $B_{protein}$ (Å^2)	22.6
Average $B_{water\ molecules}$ (Å^2)	33.8
Average $B_{sulfate}$ (Å^2)	61.1
Rmsd bond length (Å)	0.005
Rmsd angles (deg.)	1.4
Rmsd bonded *Bs* (Å^2)	1.8

[a] Values for the outer resolution shell are given in parantheses.

[b] $R_{merge} = \Sigma_{hkl}(|I_{hkl} - \langle I_{hkl}\rangle|)/\Sigma_{hkl}\langle I_{hkl}\rangle$, where I_{hkl} is the intensity of reflection *hkl*, and $\langle I_{hkl}\rangle$ is the average intensity of multiple observations.

[c] $R_{work} = \Sigma|F_o - F_c|/\Sigma F_o$, where F_o and F_c are the observed and calculated structure factor amplitudes, respectively. R_{free} is the *R*-factor for a randomly selected 5% of reflections which were not used in the refinement.

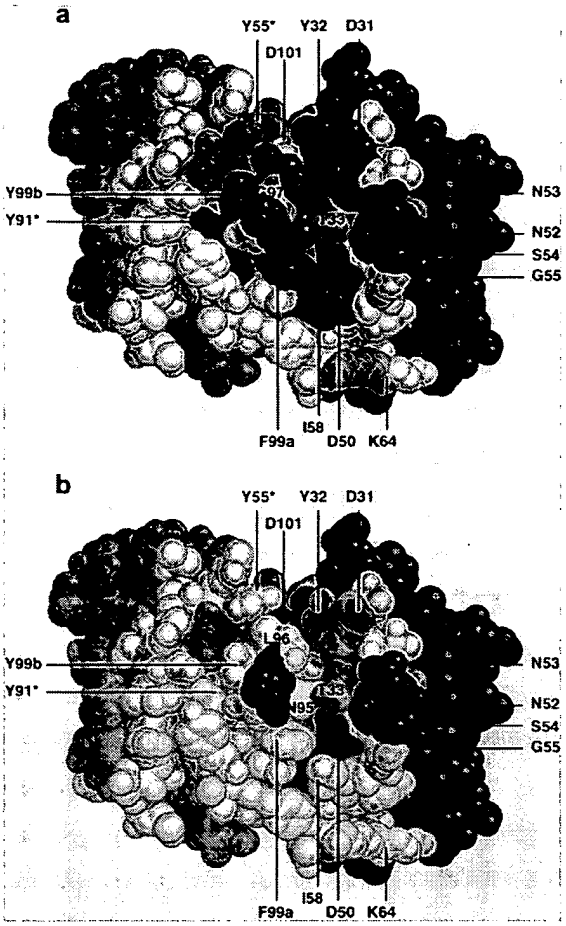


Figure 4. Mapping of the functional epitopes for binding of ErbB2-ECD onto the structure of Fab2C4. The functional epitopes defined by (a) shotgun alanine-scanning or (b) shotgun homolog-scanning are shown. Unscanned light or heavy chain residues are colored cyan or blue, respectively. Scanned residues are color-coded according to the magnitudes of $F_{wt/mut}$ values, as follows: red, >30; yellow, 10–30; grey, <10. Labeled residues with asterisks (*) indicate light chain residues. The solvent-exposed Cγ2 group of *h*T33 is located directly over a hydrogen bond network involving *h*T33 and several buried side-chains (Figure 5). Data are also shown graphically in Figure 2 and were obtained from Tables 3 and 4. The Fab2C4 structure is shown in CPK representation. This Figure and Figure 5 were generated using PyMOL (http://pymol.sourceforge.net).

domains, as the rmsd between the two structures decreases to 0.7 Å or 0.9 Å when the superposition is performed using only the constant domains or the variable domain frameworks, respectively. All residues in the CDRs of Fab2C4 are well ordered, with the exception of *h*L100. The disorder in this region appears to be correlated with disorder around the immediately adjacent CDR-H1 residue *h*Y32. In comparison with Fab4D5v4, there is a single amino acid deletion in the sequence of Fab2C4 that occurs in CDR-H3 and is accommodated by a completely altered backbone trajectory. Conservation of a hydrophobic patch in this region is maintained by the aromatic ring of *h*F105 in Fab2C4 lying in the same position as *h*W99 in Fab4D5v4, despite the fact that these residues are at opposite ends of CDR-H3. The only difference in the light chain frameworks occurs at position 66, where an arginine in Fab4D5v4 is substituted by a glycine in Fab2C4, causing the polypeptide backbone to undergo a significant rearrangement.

Discussion

Antibody affinity and specificity is predominantly dictated by the six CDR loops that together form the antigen-binding site. With the exception of CDR-H3, the conformations of the CDR main-chains do not vary greatly within different antibodies, and they can be classified into a limited number of "canonical structures".[14] Thus, the major determinants of antibody specificity and affinity are the CDR side-chains. Side-chains can be classified as buried or solvent exposed on the basis of their solvent accessible surface area, and these classifications have implications for the

Table 7. Buried Fab2C4 CDR residues and contacts

Residue	Contacts
*l*A25	*l*K24, *l*S26, *l*Q27, *l*V29
*l*V29	*l*A25, *l*Q27, *l*D28, *l*S30, *l*V33
*l*G32	*l*V29, ***l*Y91**, *l*Y92
*l*V33	*l*V29, *l*G32, *l*A34, ***l*Q89**, ***l*Y91**
*l*A34	*l*V33, ***l*Q89**, ***l*Y91**
*l*A51	*l*I31, *l*V33, *l*S50, *l*S52
***l*Q89**	*l*A34, *l*Q90, *l*Y96, *l*T97, ***h*H99a**, ***h*Y99b**, ***h*F100**
*l*Q90	*l*G32, ***l*Q89**, ***l*Y91**, *l*Y92, *l*I93, *l*Y96, *l*T97
***l*Y91**	*l*I31, *l*G32, *l*V33, *l*A34, *l*S50, *l*Q90, *l*Y92, *l*Y96, *h*S99, ***h*F99a**, ***h*Y99b**
***h*M34**	***h*T33**, ***h*D35**, ***h*V51**
***h*D35**	*h*M34, ***h*D50**, *h*N95, ***h*F99a**, ***h*F100**
***h*D50**	***h*T33**, *h*M34, ***h*D35**, ***h*V51**, ***h*I58**, *h*N95, ***h*F99a**
*h*V51	***h*M34**, ***h*D50**, ***h*N52**, ***h*G55**, *h*G56
***h*P52a**	*h*T30, ***h*Y32**, ***h*T33**, *h*V51, *h*N53
*h*F63	*h*N60, *h*R62, ***h*K64**
*h*N95	***h*T33**, ***h*D35**, ***h*D50**, ***h*L96**, *h*G97, *h*S99, ***h*F99a**, ***h*Y99b**, ***h*F100**
***h*F100**	***l*Q89**, ***h*D35**, *h*N95, ***h*F99a**, ***h*Y99b**, *h*D101

Buried CDR residues are listed in the left column. The right column lists CDR residues that contact each buried residue. A buried residue is defined as one with a solvent-accessible surface area of 10% or less than that of an identical residue in a Gly-X-Gly tripeptide. Two residues are defined as being in contact if at least one side-chain atom of the buried residue is within 4 Å of the other residue. Bold text denotes residues with $F_{wt/Ala} > 10$, indicating that alanine substitutions at these positions greatly reduce the binding affinity of Fab2C4 for the antigen ErbB2-ECD. Only residues subjected to shotgun scanning are listed.

functional role of each CDR side-chain in antigen binding.

Solvent-exposed side-chains cover the surface of the antigen-binding site where they can potentially make direct binding contacts with the antigen. Solvent-exposed side-chains that become buried upon antigen binding are components of the "structural-binding epitope", and within this group, those that make energetic contributions to the binding interaction constitute the "functional-binding epitope".[8,15] In contrast, buried side-chains are unable to make significant direct contact with antigen, but they can still contribute to antigen binding by acting as "scaffolding" residues that pack against residues in the functional epitope, and in so doing, maintain the structural integrity of the antigen-binding site.[16]

Accurate definition of the structural-binding epitope can only be achieved with an antibody–antigen co-crystal structure which provides a three-dimensional view of the binding interface. However, as we show here, considerable insights into antibody function can be gained with comprehensive CDR mutagenesis data in conjunction with the structure of an unbound antigen-binding site. Such a database can reveal both the functional epitope that makes energetically favorable contacts with antigen, and also, the buried scaffolding side-chains that serve to maintain the functional epitope in a binding-competent conformation.

The functional-binding epitope of Fab2C4

When the shotgun alanine-scan results were mapped onto the X-ray crystal structure of Fab2C4 (Figure 4(a)), they revealed a functional epitope comprising two solvent-exposed "ridges". One ridge is composed of residues *h*I58, *h*P98, *h*F99a, and *h*Y99b while the other contains residues *h*D31, *h*Y32, *h*N52, *h*N53, and *h*S54. These residues are all in the heavy chain, suggesting that much of the binding energy is derived from heavy chain interactions. Intriguingly, the most important functional epitope residues revealed by the shotgun homolog-scan (Figure 4(b)) constitute a subset of the residues found to be important in the shotgun alanine-scan. In particular, three residues in CDR-H2 (*h*N52, *h*N53, and *h*S54) form a small patch that was highly conserved in both scans, suggesting that this surface makes precise contacts with the antigen ErbB2-ECD. This prediction is further supported by affinity measurements with point-mutated proteins which also showed that these residues are extremely intolerant to substitutions (Table 5).

The functional epitope defined by the shotgun alanine-scan covers a surface area of 610 Å^2 in the Fab2C4 structure, with almost all of this area (608 Å^2) being confined to the heavy chain. In contrast, the functional epitope defined by the shotgun homolog-scan covers only 369 Å^2. Thus, the functional epitope can be divided into two distinct regions. One subset of the surface appears to make relatively non-specific interactions with ErbB2-ECD, while a much smaller surface appears to make more specific contacts with the antigen. Interestingly, this smaller surface, revealed as functionally important by both scans, is restricted entirely to the heavy chain; the side-chains found to be important only in the alanine-scan are located at the dimer interface between the light and heavy chains. The size of the functional epitope identified by the alanine-scan is similar to that of structural-binding epitopes revealed by X-ray crystal structures of antibody–antigen complexes.[17,18]

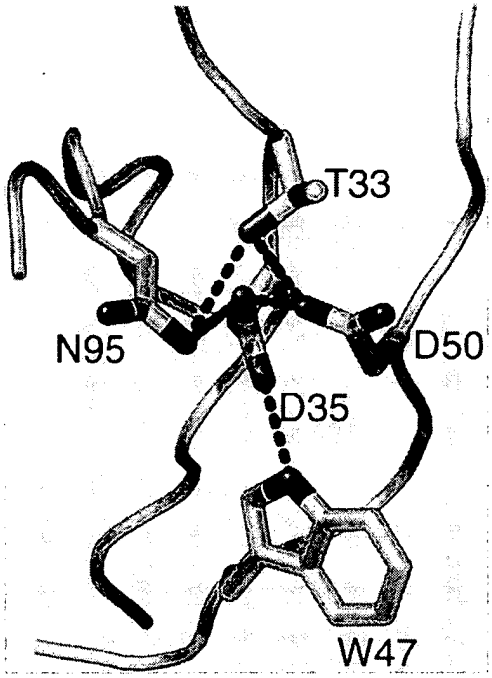


Figure 5. The buried hydrogen bond network in the heavy chain of Fab2C4. Side-chain atoms are shown as sticks colored as follows: carbon, gray; oxygen, red; nitrogen, blue. Main chains are shown as gray tubes and hydrogen bonds are shown as dashed magenta lines. The hydrogen bond network ties together residues from each of the three heavy chain CDRs, and with the exception of *h*T33, all the interacting side-chains are buried. The γ-OH group of *h*T33 is anchored by the hydrogen bonds, resulting in the Cγ2 group being solvent exposed at the base of a central depression in the Fab2C4 functional epitope (Figure 4).

The role of scaffold residues in antigen binding

Of the 61 CDR residues that were subjected to shotgun scanning, 17 can be classified as buried (Table 7), if a buried residue is defined as one with a solvent accessible surface area of 10% or less than that of an identical residue in a Gly-X-Gly tripeptide. Substitutions for eight of the nine buried light chain residues have little effect on ErbB2-ECD binding. *l*Y91 is the only buried light chain residue which was found to be intolerant to substitution in the alanine-scan (Table 3 and Figure 2(a)). However, the homolog-scan showed that substitution of *l*Y91 with phenyalanine is well tolerated, indicating that the contribution of the tyrosine side-chain to antigen-binding results from the volume occupied by the large, hydrophobic phenyl ring rather than the side-chain hydroxyl group. *l*Y91 makes contact with a number of CDR residues, including heavy chain residues that contribute significantly to antigen binding (Table 7), and thus, it appears to buttress the functional-binding epitope. Similar buttressing interactions have been shown to effect the binding affinities of other antigen–antibody interactions.[16,19]

In contrast to the light chain residues, most buried heavy chain residues appear to be involved in scaffolding interactions that are critical for antigen binding. All eight buried heavy chain residues that we scanned preferred wt over alanine, with six residues exhibiting $F_{wt/ala}$ values greater than 10 (Table 4 and Figure 2(b)). The buried heavy chain residues make numerous contacts with functionally important CDR residues (Table 7), and in so doing, apparently fix the Fab2C4 functional epitope in a productive binding conformation.

Antibody CDRs contain an unusually large proportion of buried hydrophilic residues, in comparison with other protein folds or even other regions of antibody structure.[20] Indeed, many of the buried CDR residues in Fab2C4 are hydrophilic (Table 7), and several of these side-chains participate in important scaffolding interactions that are essential for high affinity antigen binding (Figure 2(b)). In particular, it appears that the conformations of the heavy chain CDRs are maintained not only by hydrophobic contacts, but also by a hydrogen bond network involving buried CDR side-chains (Figure 5). Although the Fab2C4 crystals were grown at near neutral pH, the buried side-chains of *h*D35 and *h*D50 are oriented nearly directly towards each other and their Oδ1 atoms are in close proximity (2.4 Å distance), suggesting that one or both of these carboxylates exist in a protonated state. The hydrogen bond network also involves the side-chains of *h*N95, *h*T33 and *h*W47, and thus, it connects residues from each of the three heavy chain CDRs. The need for precise geometry and chemistry within the network is reflected in the intolerance of the interacting side-chains to substitution not only with alanine, but also with homologous side-chains (Table 4 and Figure 2(b)).

Reproducibility and reliability of shotgun scanning

Shotgun scanning uses statistical analysis of combinatorial libraries to predict the functional contributions of individual protein side-chains. For these predictions to be valid, two important conditions must be met. First, the prediction for any individual mutation should be independent of the combinatorial library used for the scan, or in other words, the predicted effect of an identical mutation in two different libraries should be the same. Second, the shotgun scanning predictions should be in good agreement with results obtained with affinity measurements for point-mutated proteins. The Fab2C4 shotgun scan results meet both criteria.

The alanine and homolog-scan data sets contained 26 overlapping $F_{wt/mut}$ values (asterisks in Tables 3 and 4), allowing us to compare the predicted functional effects for identical mutations obtained from different libraries. As shown in

Figure 3, the two scans gave remarkably similar estimates of functional importance for identical mutations. Thus, data from different libraries can be compared with each other with high confidence, as the results for individual mutations are highly reproducible and independent of the library. Furthermore, these results suggest that most side-chains in the Fab2C4 CDRs act additively and independently in the binding interaction with ErbB2-ECD, since the effects of individual mutations do not appear to be significantly influenced by mutations at other CDR positions. The shotgun scanning predictions are also in good agreement with affinity data for purified point mutants. Shotgun homolog-scanning predicted that the mutation *h*M34L should result in a modest 2.5-fold reduction in affinity (Table 4), and indeed, the introduction of this mutation into Fab2C4 reduced binding to ErbB2-ECD by about 2-fold (Table 5). In contrast, shotgun scanning identified residues *h*N52, *h*N53, and *h*S54 as critical components of a small solvent-exposed patch that was highly intolerant to substitutions. Again, affinity measurements with purified proteins confirmed this prediction, as all substitutions at these sites resulted in Fab2C4 mutants with greatly compromised antigen-binding activity (Table 5).

Thus, the results of the two shotgun scans were in good agreement with each other and with the results obtained with purified point-mutated proteins. In conjunction with the Fab2C4 crystal structure, the scanning data provided comprehensive functional maps of the Fab2C4 antigen-binding site. Together, the alanine and homolog-scanning maps allowed for the accurate definition of both the functional binding epitope, and also, the buried scaffolding side-chains that hold this epitope in a binding-competent conformation. The data should be useful in guiding the design of new phage-displayed libraries for further optimization of the binding interaction between Fab2C4 and ErbB2-ECD. Furthermore, the general shotgun scanning technology should facilitate the rapid analysis of many other protein–protein interactions.

Materials and Methods

Oligonucleotides

DNA degeneracies are represented in the IUB code (K = G/T, M = A/C, R = A/G, S = G/C, W = A/T, Y = C/T). Degenerate codons are shown in bold text. The following mutagenic oligonucleotides were used for library constructions:

H1-A1: GCA GCT TCT GGC TTC **RCT** TTC **RCT GMT KMT RCT** ATG GAC TGG GTC CGT

H1-A2: GCA GCT TCT GGC TTC ACC TTC ACC GAC TAT ACC ATG **GMT** TGG GTC CGT CAG GCC

H2-A1: CTG GAA TGG GTT GCA **GMT GYT RMC** CCT **RMC KCC** GGC GGC TCT **RYT** TAT **RMC SMA** CGC TTC AAG GGC CGT

H2-A2: CTG GAA TGG GTT GCA GAT GTT AAT **SCA** AAC AGT **GST GST KCC** ATC **KMT** AAC CAG **SST KYT RMA GST** CGT TTC ACT CTG AGT

H3-A1: TAT TAT TGT GCT CGT **RMC SYT** GGA **SCA KCC** TTC TAC TTT GAC TAC

H3-A2: TAT TAT TGT GCT CGT AAC CTG **GST** CCC TCT **KYT KMT KYT GMT KMT** TGG GGT CAA GGA ACC

H1-H: GCA GCT TCT GGC TTC **ASC** TTC **ASC GAM TWC ASC MTG GAM** TGG GTC CGT CAG GCC

H2-H: GGC CTG GAA TGG GTT GCA **GAM RTT RAC SCA RAC KCC GST GST KCC RTT TWC RAC SAA** CGC TTC AAG GCC CGT

H3-H: TAT TAT TGT GCT CGT **RAC MTC GST SCA KCC TWC TWC TWC GAM TWC** TGG GGT CAA GGA ACC

L1-A1: ACC TGC AAG GCC AGT **SMA GMT** GTG **KCC RYT GST** GTC GCC TGG TAT CAA

L1-A2: GTC ACC ATC ACC TGC **RMA GST KCC** CAG GAT **GYT** TCT ATT GGT **GYT GST** TGG TAT CAA CAG AAA CCA

L2-A1: AAA CTA CTG ATT TAC **KCC** GCT **KCC KMT** CGA **KMT** ACT GGA GTC CCT TCT

L2-A2: AAA CTA CTG ATT TAC TCG **GST** TCC TAC **SST** TAC **RCT** GGA GTC CCT TCT CGC

L3-A1: TAT TAC TGT CAA CAA **KMT KMT RYT KMT** CCT **KMT** ACG TTT GGA CAG GGT

L3-A2: GCA ACT TAT TAC TGT **SMA SMA** TAT TAT ATT TAT **SCA** TAC **RCT** TTT GGA CAG GGT ACC

L1-H: GTC ACC ATC ACC TGC **ARG KCC KCC SAA GAM RTT KCC RTT GST RTT KCC** TGG TAT CAA CAG AAA CCA

L2-H: AAA CTA CTG ATT TAC **KCC KCC KCC TWC ARG TWC ASC** GGA GTC CCT TCT CGC

L3-H: GCA ACT TAT TAC TGT **SAA SAA TWC TWC RTT TWC SCA TWC ASC** TTT GGA CAG GGT ACC

Construction of shotgun scanning libraries

To display Fab2C4 on the surface of M13 bacteriophage, we modified a previously described phagemid (pS1602) designed for the display of human growth fused to the C-terminal domain of the M13 gene-3 minor coat protein (cP3).[21] Standard molecular biology techniques were used to replace the fragment of pS1602 encoding human growth hormone with a DNA fragment encoding the light and heavy chains of Fab2C4. The resulting phagemid (pC2C4) contained a bicistronic gene under the control of the IPTG-inducible P_{tac} promoter.[22] The first open reading frame encoded a polypeptide consisting of the maltose binding protein secretion signal, followed by an epitope tag (amino acid sequence: MADPNRFRGKDLGG),[23] followed by the Fab2C4 light chain. The second open reading frame encoded a polypeptide consisting of the stII secretion signal,[24] followed by the Fab2C4 heavy chain, followed by cP3. Expression in *E. coli* resulted in the periplasmic secretion of free Fab2C4 light chain and Fab2C4 heavy chain fused to cP3, and the two chains spontaneously associated to assemble functional Fab2C4. *E. coli* cultures harboring pC2C4 were coinfected with M13-VCS helper phage (Stratagene) and grown at 37 °C without IPTG induction, resulting in the production of M13 bacteriophage displaying Fab2C4 in a monovalent format.

Libraries were constructed using previously described methods[25] with appropriately designed "stop template" versions of pC2C4. For each library, we used a stop template which contained TAA stop codons within each of the CDRs to be mutated. The stop template was used as the template for the Kunkel mutagenesis method[26] with mutagenic oligonucleotides (see above) designed to simultaneously repair the stop codons and introduce mutations at the desired sites. The libraries are described in Table 2.

Library sorting and binding assays

NUNC 96-well Maxisorp immunoplates were coated overnight at 4 °C with capture target (either ErbB2-ECD or an anti-tag antibody at 5 μg/ml) and blocked for two hours with bovine serum albumin (BSA) (Sigma). Phage from the libraries described above were propagated in *E. coli* XL1-blue with the addition of M13-VCS helper phage. After overnight growth at 37 °C, phage were concentrated by precipitation with PEG/NaCl and resuspended in phosphate-buffered saline (PBS), 0.5% (w/v) BSA, 0.1% (v/v) Tween 20 (Sigma), as described previously.[25] Phage solutions (10^{13} phage/ml) were added to the coated immunoplates. Following a two-hour incubation to allow for phage binding, the plates were washed 12 times with PBS, 0.05% Tween 20. Bound phage were eluted with 0.1 M HCl for ten minutes and the eluant was neutralized with 1.0 M Tris base. Eluted phage were amplified in *E. coli* XL1-blue and used for further rounds of selection.

Individual clones from each round of selection were grown in a 96-well format in 500 μl of 2YT broth supplemented with carbenicillin and M13-VCS, and the culture supernatants were used directly in phage ELISAs[25] to detect phage-displayed Fab2C4 variants that bound to either ErbB2-ECD or anti-tag antibody. After one round of selection for anti-tag binding or two rounds of selection for ErbB2-ECD binding, greater than 50% of the clones exhibited positive phage ELISA signals at least 2-fold greater than signals on control plates coated with BSA. These positive clones were subjected to DNA sequence analysis.

DNA sequencing and analysis

Culture supernatants containing phage particles were used as templates for PCRs that amplified DNA fragments containing the light and heavy chain genes. The PCR primers were designed to add M13(-21) and M13R universal sequencing primers at either end of the amplified fragment, thus facilitating the use of these primers in sequencing reactions. Amplified DNA fragments were sequenced using Big-Dye terminator sequencing reactions which were analyzed on an ABI Prism 3700 96-capillary DNA analyzer (PE Biosystems, Foster City, CA). All reactions were performed in a 96-well format.

The sequences were analyzed with the program SGCOUNT as described previously.[11] SGCOUNT aligned each DNA sequence against the wild-type DNA sequence by using a Needleman–Wunch pairwise alignment algorithm, translated each aligned sequence of acceptable quality, and tabulated the occurrence of each natural amino acid at each position. For the ErbB2-ECD binding selection, the number of analyzed clones are indicated in parenthesis following the name of each library: HAa (178), HAb (67), LAa (167), LAb (85), HHa (89), HHb (64), LH (65). For the anti-tag antibody-binding selection, the following number of clones were analyzed for each library: HAa (94), HAb (91), LAa (183), Lab (72), HHa (73), HHb (73), LH (84).

ELISA for measuring the binding of Fab2C4 mutants to ErbB2-ECD

Fab2C4 and mutant proteins were produced using a previously described expression vector with an alkaline phosphatase promoter.[27] Fab2C4 mutants were constructed using the Kunkel site-directed mutagenesis method.[26] For protein production, expression plasmids were transformed into *E. coli* 16C9, cultures were grown in modified AP5 medium, and Fab proteins were purified using protein G-sepharose CL-4B (Amersham Pharmacia), as described previously.[27] Fab proteins were buffer exchanged into 10 mM sodium succinate, 140 mM NaCl (pH 6.0), and concentrated using a Centricon-10 (Amicon). Protein concentrations were determined by quantitative amino acid analysis.

The binding activities of Fab2C4 and mutants were measured with an ELISA. NUNC 96-well maxisorp immunoplates were coated overnight at 4 °C with ErbB2-ECD (1 μg/ml in 50 mM carbonate buffer (pH 9.6)). The plates were blocked for one hour at room temperature with ELISA diluent buffer (PBS, 0.5% BSA, 0.05% Tween 20). Serial dilutions of Fab protein were incubated on the ErbB2-ECD-coated plates for two hours at room temperature, and the plates were washed with PBS, 0.05% Tween 20. Bound Fab proteins were detected with biotinylated murine anti-human kappa chain antibody followed by streptavidin/horseradish peroxidase conjugate (Sigma), using 3,3′,5,5′-tetramethyl benzidine (TMB) as substrate (Kirkegaard and Perry Laboratories, Gaithersburg, MD). Titration curves were fit with a four-parameter non-linear regression curve-fitting program (KaleidaGraph, Synergy Software) to determine the EC_{50} values, the Fab concentrations corresponding to half-maximal binding signals. The fold reduction in ErbB2-ECD binding activity due to each point mutation was determined by dividing the EC_{50} for the Fab2C4 point mutant by the EC_{50} for wild-type Fab2C4 (Table 5).

Purification of Fab2C4 protein for X-ray crystallography

Fab 2C4 was produced in *E. coli* 16C9 using a previously described expression vector with an alkaline phosphatase promoter.[27] Cells were pelleted by centrifugation, resuspended in PBS (5 ml/g of cell paste), and disrupted by three passes through a Microfluidizer HC8000 (7000 PSI). The suspension was adjusted to 50 mM magnesium sulfate and 0.2% (v/v) polyethylene imine, stirred for 30 minutes, and centrifuged for one hour in a Sorval RC3B centrifuge (4500 rpm) to remove solids. The supernatant was decanted and sterile filtered (0.22 μm). A Gammabind + protein G Sepharose column (Amersham Pharmacia) was equilibrated in PBS, the supernatant was loaded at 40 ml/ml of resin, and eluted in 0.1 M acetic acid, 25 mM NaCl (pH 2.8). The eluant was adjusted to pH 6.0 with 2.0 M Tris base, mixed with an equal volume of 1.0 M sodium sulfate, sterile filtered (0.22 μm), and loaded on a J. T. Baker Hi-Propyl column (Mallinckrodt-Baker, Phillipsburg, NJ) at 5 mg/ml of resin. Fab2C4 protein was eluted with a 20-column volume gradient from 0.5 M sodium sulfate, 10 mM succinate (pH 5.0) to 10 mM succinate (pH 5.0). The

Hi-propyl pool was diluted to a conductivity of 2.5 mS with purified water and loaded on a J. T. Baker ABX column (Mallinckrodt-Baker, Phillipsburg, NJ) at 1.0 mg/ml of resin. The ABX column was eluted with a 20 column volume gradient from 50 mM Hepes (pH 7.1) to 100 mM NaCl, 50 mM Hepes (pH 7.1). The resulting pool was concentrated in an Amicon stirred cell with a YM10 membrane (Millipore, Bedford, MA) and formulated in 10 mM succinate (pH 5.0) on a sephadex G-25 column (Amersham Pharmacia) (0.1 ml/ml of resin).

Crystallization and data collection

Prior to crystallization experiments, Fab2C4 protein was dialyzed against deionized water, then concentrated to 22.5 mg/ml using a Centricon 10 microconcentrator (Millipore, Bedford, MA). Crystallization was performed with the sitting drop vapor diffusion method and used 2 μl protein solution and 2 μl precipitant (35–45% saturated $(NH_4)_2SO_4$, 0.1 M Tris–HCl (pH 8.0)). After 24 hours, the drops were microseeded with 1.0 μl of serial dilutions of a fresh seed stock solution. Thin, plate-like crystals appeared within hours, and grew to a maximum size of 0.1 mm × 0.05 mm × 0.01 mm within one week. Crystals were harvested into a cryo-protectant solution consisting of 45% saturated $(NH_4)_2SO_4$, 0.1 M Tris–HCl, 20% (v/v) glycerol (pH 8.0), and immediately flash cooled in liquid nitrogen. Initially, a data set complete to 2.7 Å was collected using a Mar image plate detector mounted on a rotating anode X-ray source operating at a wavelength of 1.5418 Å. This data set was used for the molecular replacement calculations and for the initial refinement steps. Later, a separate 1.8 Å data set was collected from the same specimen at the Advanced Light Source, Beamline 5.0.2 at a wavelength of 1.1 Å. The data were integrated using the program MOSFLM[28,29] and scaled using the program SCALA.[30,31] Intensities were converted to amplitudes by the method of French & Wilson[32] using the program TRUNCATE, and Free *R* flags were assigned using the program UNIQUE.[31] Data statistics are reported in Table 6.

Structure determination and refinement

The structure was determined by molecular replacement using the coordinates of Fab4D5 version 4 (PDB code 1fvd)[12] and the program AMORE.[31,33] All data between 10 Å and 3 Å were used. The highest peak in the rotation function search had a correlation coefficient (CC) of 0.28, while the next highest peak had a CC of 0.18. The correct solution was readily identified in the translation function, with a CC of 0.38 (next highest 0.20) and an *R*-factor of 49% (next best 54.2%). Rigid body refinement further improved the model, producing a CC of 47.3%, and an *R*-factor of 47.2%. The single molecule in the asymmetric unit of these crystals results in a solvent content of 42% and a corresponding Matthews coefficient of 2.1 $Å^3$/Da. At this point, non-identical residues between Fab4D5 and Fab2C4 were removed from the model and the model was subjected to rigid body refinement using the program CNX[34] (MSI, San Diego, CA), and all data between 6.0 Å and 2.7 Å, treating the variable and constant domains of the heavy and light chains as independent groups. This reduced the *R*-factor to 46.9% (R_{free} = 46.6%). Torsion angle molecular dynamics, followed by Powell minimization and grouped (main-chain and side-chain) *B*-factor refinement reduced the *R*-factor against 8–2.7 Å data to 32.3% (R_{free} = 42%). Automatic refinement continued using the high resolution data, this time using the Cartesian molecular dynamics simulated annealing protocol in CNX. The *R*-factor at this point was 27.1% (R_{free} = 30.5%). Inspection of SigmaA weighted maps[32,35] clearly indicated the conformations of the omitted residues, and 12 cycles of automatic refinement, followed by manual rebuilding, resulted in the final model, which consists of residues 1–214 of the light chain, residues 1–223 of the heavy chain, 382 water molecules, and two sulfate ions. The final *R*-factor was 19.7% (R_{free} = 23.0%). The refinement statistics are shown in Table 6.

Protein Data Bank accession number

The coordinates and structure factors of Fab2C4 have been deposited with the Research Collaboratory for Structural Bioinformatics PDB[36] under accession number 1L7I.

Acknowledgments

We thank Alan Zhong for DNA sequencing and the Genentech DNA synthesis group for oligonucleotides. We thank the staff of the Macromolecular Crystallography Facility at the Advanced Light Source (Berkeley, CA) for assistance with the high-resolution X-ray data set. The Advanced Light Source is supported by the Director, Office of Science, Office of Basic Energy Sciences, Materials Sciences Division, of the US Department of Energy under Contract No. DE-AC03-76SF00098 at Lawrence Berkeley National Laboratory.

References

1. Hudson, P. J. (1999). Recombinant antibody constructs in cancer therapy. *Curr. Opin. Immunol.* **11**, 547–557.
2. Dall'Acqua, W. & Carter, P. (1998). Antibody engineering. *Curr. Opin. Struct. Biol.* **8**, 443–450.
3. Chothia, C., Lesk, A. M., Tramontano, A., Levitt, M., Smith-Gill, S. J., Air, G. *et al.* (1989). Conformations of immunoglobulin hypervariable regions. *Nature*, **342**, 877–883.
4. Tramontano, A., Chothia, C. & Lesk, A. M. (1990). Framework residue 71 is a major determinant of the position and conformation of the second hypervariable region in the V_H domains of immunoglobulins. *J. Mol. Biol.* **215**, 175–182.
5. Padlan, E. A., Silverton, E. W., Sheriff, S., Cohen, G. H., Smith-Gill, S. J. & Davies, D. R. (1989). Structure of an antibody–antigen complex: crystal structure of the HyHEL-10 Fab-lysozyme complex. *Proc. Natl Acad. Sci. USA*, **86**, 5938–5942.
6. Bentley, G. A., Boulot, G., Riottot, M. M. & Poljak, R. J. (1990). Three-dimensional structure of an idiotope–anti-idiotope complex. *Nature*, **348**, 254–257.
7. Wells, J. A. (1991). Systematic mutational analyses of protein–protein interfaces. *Methods Enzymol.* **202**, 390–411.
8. Clackson, T. & Wells, J. A. (1995). A hot spot of binding energy in a hormone–receptor interface. *Science*, **267**, 383–386.

9. Kelley, R. F. & O'Connell, M. P. (1993). Thermodynamic analysis of an antibody functional epitope. *Biochemistry*, **32**, 6828–6835.
10. Chen, G., Dubrawsky, I., Mendez, P., Georgiou, G. & Iverson, B. L. (1999). *In vitro* scanning mutagenesis of all the specificity determining residues in an antibody binding site. *Protein Eng.* **12**, 349–356.
11. Weiss, G. A., Watanabe, C. K., Zhong, A., Goddard, A. & Sidhu, S. S. (2000). Rapid mapping of protein functional epitopes by combinatorial alanine scanning. *Proc. Natl Acad. Sci. USA*, **97**, 8950–8954.
12. Eigenbrot, C., Randal, M., Presta, L., Carter, P. & Kossiakoff, A. A. (1993). X-ray structures of the antigen-binding domains from three variants of humanized anti-p185HER2 antibody 4D5 and comparison with molecular modeling. *J. Mol. Biol.* **229**, 969–995.
13. Fendly, B. M., Winget, M., Hudziak, R. M., Lipari, M. T., Napier, M. A. & Ullrich, A. (1990). Characterization of murine monoclonal antibodies reactive to either the human epidermal growth factor or HER2/*neu* gene product. *Cancer Res.* **50**, 1550–1558.
14. Chothia, C., Lesk, A. M., Tramontano, A., Levitt, M., Smith-Gill, S. J., Air, G. *et al.* (1989). Conformations of immunoglobulin hypervariable regions. *Nature*, **342**, 877–883.
15. Clackson, T., Ultsch, M. H., Wells, J. A. & de Vos, A. M. (1998). Structural and functional analysis of the 1:1 growth hormone:receptor complex reveals the molecular basis for receptor affinity. *J. Mol. Biol.* **277**, 1111–1128.
16. Schildbach, J. F., Near, R. I., Bruccoleri, R. E., Haber, E., Jeffrey, P. D., Novotny, J. *et al.* (1993). Modulation of antibody affinity by a non-contact residue. *Protein Sci.* **2**, 206–214.
17. Janin, J. & Chothia, C. (1990). The structure of protein–protein recognition sites. *J. Biol. Chem.* **265**, 16027–16030.
18. Davies, D. R., Sheriff, S. & Padlan, E. A. (1988). Antibody–antigen complexes. *J. Biol. Chem.* **263**, 10541–10544.
19. Hawkins, R. E., Russell, S. J., Baier, M. & Winter, G. (1993). The contribution of contact and non-contact residues of antibody in the affinity of binding to antigen. *J. Mol. Biol.* **234**, 958–964.
20. Padlan, E. A. (1990). On the nature of antibody combining sites: unusual structural features that may confer on these sites an enhanced capacity for binding ligands. *Proteins: Struct. Funct. Genet.* **7**, 112–124.
21. Sidhu, S. S. (2000). Phage display in pharmaceutical biotechnology. *Curr. Opin. Biotechnol.* **11**, 610–616.
22. Amman, E. & Brosius, J. (1985). "ATG vectors" for regulated high-level expression of cloned genes in *Escherichia coli*. *Gene*, **40**, 183–190.
23. Lasky, L. A. & Dowbenko, D. J. (1984). DNA sequence analysis of the type-common glycoprotein-D genes of herpes simplex virus types 1 and 2. *DNA*, **3**, 23–29.
24. Lowman, H. B., Bass, S. H., Simpson, N. & Wells, J. A. (1991). Selecting high-affinity binding proteins by monovalent phage display. *Biochemistry*, **30**, 10832–10838.
25. Sidhu, S. S., Lowman, H. B., Cunningham, B. C. & Wells, J. A. (2000). Phage display for selection of novel binding peptides. *Methods Enzymol.* **328**, 333–363.
26. Kunkel, T. A., Roberts, J. D. & Zakour, R. A. (1987). Rapid and efficient site-specific mutagenesis without phenotypic selection. *Methods Enzymol.* **154**, 367–382.
27. Presta, L. G., Chen, H., O'Connor, S. J., Chisholm, V., Meng, Y. G., Krummen, L. *et al.* (1997). Humanization of an anti-vascular endothelial growth factor monoclonal antibody for the therapy of solid tumors and other disorders. *Cancer Res.* **57**, 4593–4599.
28. Powell, H. R. (1999). The Rossman Fourier autoindexing algorithm in MOSFLM. *Acta Crystallog. sect. D*, **55**, 1690–1695.
29. Leslie, A. G. W. (1992). Recent changes to the MOSFLM package for processing film and image plate data. *Joint CCPR ESF-EAMCB Newsletter Protein Crystallog.* **26**.
30. Evans, P. R. (1993). Data reduction. *Proceedings of CCP4 Study Weekend*
31. Collaborative Computational Project No. 4 (1994). The CCP4 suite: programs for protein crystallography. *Acta Crystallog. sect. D*, **50**, 760–763.
32. French, S. & Wilson, K. (1978). On the treatment of negative intensity observations. *Acta Crystallog. sect. A*, **34**, 517–525.
33. Navaza, J. (1994). AmoRe: an automated package for molecular replacement. *Acta Crystallog. sect. D*, **50**, 157–163.
34. Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W. *et al.* (1998). Crystallography and NMR system: a new software suite for macromolecular structure determination. *Acta Crystallog. sect. D*, **54**, 905–921.
35. Read, R. J. (1986). Improved Fourier coefficients for maps using phases from partial structures with errors. *Acta. Crystallog. sect. D*, **42**, 140–149.
36. Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H. *et al.* (2000). The Protein Data Bank. *Nucl. Acids Res.* **28**, 235–242.
37. Kabat, E. A., Wu, T. T., Redi-Miller, M., Perry, H. M. & Gottesman, K. S. (1987). *Sequences of Proteins of Immunological Interest*, 4th edit., National Institutes of Health, Bethesda, MD.

Edited by I. Wilson

(Received 14 January 2002; received in revised form 19 March 2002; accepted 22 March 2002)

British Journal of Cancer (2000) 83(2), 252–260
© 2000 Cancer Research Campaign
doi: 10.1054/ bjoc.2000.1226, available online at http://www.idealibrary.com on IDEAL®

Human anti-CD30 recombinant antibodies by guided phage antibody selection using cell panning

A Klimka[1,2], B Matthey[1], RC Roovers[2], S Barth[1], J-W Arends[2], A Engert [1] and HR Hoogenboom[2]

[1]Laboratory of Immunotherapy, Department of Internal Medicine I, University Hospital Cologne, Joseph Stelzmann Str. 9, 50931 Cologne, Germany; [2]Department of Pathology, Maastricht University, PO Box 5616, 6200 MD Maastricht, the Netherlands

Summary In various clinical studies, Hodgkin's patients have been treated with anti-CD30 immunotherapeutic agents and have shown promising responses. One of the problems that appeared from these studies is the development of an immune response against the non-human therapeutics, which limits repeated administration and reduces efficacy. We have set out to make a recombinant, human anti-CD30 single-chain variable fragment (scFv) antibody, which may serve as a targeting moiety with reduced immunogenicity and more rapid tumour penetration in similar clinical applications. Rather than selecting a naive phage antibody library on recombinant CD30 antigen, we used guided selection of a murine antibody in combination with panning on the CD30-positive cell line L540. The murine monoclonal antibody Ki-4 was chosen as starting antibody, because it inhibits the shedding of the extracellular part of the CD30 antigen. This makes the antibody better suited for CD30-targeting than most other anti-CD30 antibodies. We have previously isolated the murine Ki-4 scFv by selecting a mini-library of hybridoma-derived phage scFv-antibodies via panning on L540 cells. Here, we report that phage display technology was successfully used to obtain a human Ki-4 scFv version by guided selection. The murine variable heavy (VH) and light (VL) chain genes of the Ki-4 scFv were sequentially replaced by human V gene repertoires, while retaining only the major determinant for epitope-specificity: the heavy-chain complementarity determining region 3 (CDR3) of murine Ki-4. After two rounds of chain shuffling and selection by panning on L540 cells, a fully human anti-CD30 scFv was selected. It competes with the parental monoclonal antibody Ki-4 for binding to CD30, inhibits the shedding of the extracellular part of the CD30 receptor from L540 cells and is thus a promising candidate for the generation of anti-CD30 immunotherapeutics. © 2000 Cancer Research Campaign

Keywords: CD30; phage display; chain shuffling; human antibody; guided selection

Although monoclonal antibodies (moab) raised by hybridoma technology (Köhler and Milstein, 1975) have been demonstrated to be very useful in research and diagnosis, they are somewhat problematic as binding moieties in immunotherapeutic agents for the treatment of tumours. Apart from their relatively large size (150 kDa), which makes it difficult to penetrate into solid tumours, these non-human antibodies generate an immune response resulting in serious side-effects such as serum sickness or anaphylactic shock, which prevent long-term treatment of cancer patients (Shawler et al, 1985). It is also documented that this human anti-mouse antibody (HAMA) response causes a rapid blood clearance of these reagents, which diminishes their efficacy (Khazaeli et al, 1994).

To circumvent these problems, two strategies have been followed. First, a reduction of the molecular size of the binding moiety using Fab fragments or even just the variable fragments of an antibody as a single-chain variable fragment (scFv) has significantly reduced the target surface for an immune response and thus the immunogenicity. Secondly, the use of humanized proteins like chimaeric or CDR-grafted, or even fully human antibodies or antibody fragments, has been demonstrated to reduce their immunogenicity (Meredith et al, 1993).

Received 6 December 1999
Revised 6 March 2000
Accepted 10 March 2000

Correspondence to: HR Hoogenboom

Because of technical problems and difficulties in retrieving suitable human donors, it is complicated to raise human hybridomas by conventional techniques. However, the progress in molecular biology has offered different ways to evade this restriction. One possibility is the use of transgenic mice carrying human immunoglobulin genes. These mice can be used to generate hybridomas secreting human antibodies (Brüggemann and Neuberger, 1996). Another way is the use of human V-gene libraries expressed and displayed on phage and selection of antigen-specific antibodies therefrom (Hoogenboom, 1997; Winter et al, 1994). These libraries can be derived from immunized or non-immunized donors or even generated synthetically (Hoogenboom et al, 1998). Indeed, from very large phage libraries, high-affinity antibodies to many different target antigens can be selected (Hoogenboom, 1997). This *in vitro* selection procedure is subjected to a series of biases introduced by library preparation, selection conditions and the screening protocol. Strong biases in selected populations can arise, in particular when selecting on complex antigenic targets (Hoogenboom et al, 1999; Persic et al, 1999).

Therefore, it sometimes remains difficult to retrieve antibodies with desired properties like recognition of a unique epitope, induction of a post-binding signal transduction or internalization upon binding to a cell-surface receptor on the target cell (McCall et al, 1998). Hybridoma-derived antibodies with such characteristics are sometimes available, and may be converted to human versions using a method termed 'guided selection' (Jespers et al, 1994): by

two consecutive chain-shuffling procedures, the rodent antibody domains are swapped for human domains, using phage display technology (library construction and selection on antigen) to retrieve the best-matching partner. Our goal was to obtain a fully human antibody from the well-characterized murine moab Ki-4, which recognizes the CD30 receptor.

CD30 was originally identified by Schwab et al (1982) as the antigen abundantly expressed on Hodgkin–Reed Sternberg cells (H–RS) in primary Hodgkin's lymphoma and recognized by the first anti-CD30 moab Ki-1. Expression of CD30 in high copy numbers on the cell surface has also been reported for a subset of non-Hodgkin lymphomas (NHL), virally transformed B- and T-cell lines, a subform of large-cell anaplastic lymphoma (CD30$^+$-LCAL), embryonal carcinomas, malignant melanomas and mesenchymal tumours (Gruss and Dower, 1995). The CD30 receptor is therefore a useful clinical and pathological tumour marker for these diseases and a good target for immunotherapy.

Here we report the synthesis of a human anti-CD30 scFv (hAK30) on the basis of the murine anti-CD30 moab Ki-4. Murine moab Ki-4, which shows no detectable cross-reactivity with vital human organs, has successfully been used as part of a chemically linked ricin A immunotoxin in vivo (Schnell et al, 1995) and also as a scFv in a *Pseudomonas* exotoxin A-based recombinant immunotoxin in vitro (Klimka et al, 1999). Therefore, we exchanged the murine variable heavy (VH) and light (VL) chain genes with human counterparts with respective selections on the CD30-positive Hodgkin cell line L540. This strategy allowed the construction of a fully human anti-CD30 scFv (hAK30) with the same binding specificity as moab Ki-4, and in which only the VH(CDR3) and framework 4 sequences are derived from the parental antibody. This scFv may serve as a useful building block for the synthesis and engineering of different fusion proteins, such as scFv coupled to toxins, enzymes, or, in connection with other targeting molecules, as bispecific agents (Huston et al, 1993). It is a promising candidate to use as immunotherapeutic agent for the treatment of CD30-positive malignancies.

MATERIAL AND METHODS

Cell lines

The Hodgkin-derived cell line L540 (Diehl et al, 1981) and the hybridoma cell lines BW702 (Bosslet et al, 1989), Ki-3, Ki-4, Ki-6, Ki-7 (Horn-Lohrens et al, 1995) and BerH2 (Schwarting et al, 1989) were maintained in RPMI 1640 medium (GIBCO-BRL, Rockville, MD, USA) supplemented with 10% (v/v) FCS, 100 µg ml^{-1} streptomycin, 200 units ml^{-1} penicillin and 2 mM L-glutamine (10% FCS-medium). All cells were cultured at 37°C in a 5% CO_2 humidified atmosphere.

Bacterial strains and plasmids

E.coli XL1-Blue (supE44, hsdR17, recA1, endA1, gyr, A46, thi, relA1, lacF′. proA$^+$ B$^+$ lacI^q, lacZ_M15, Tn10(tet^r)] were obtained from Stratagene (La Jolla, CA, USA). *E.coli* TG1 (K12_(lac-pro), supE, thi, hsdD5/F′traD36, proA$^+$ B$^+$, lacI^q, lacZ_M15) and *E.coli* HB2151 (K12_(lac-pro), ara, nal^r, thi/F′, proA$^+$ B$^+$, lacI^q, lacZ_M15) were purchased from Pharmacia (Uppsala, Sweden). The phagemid vector pCANTAB6 (McCafferty et al, 1994) is used for N-terminal fusion of scFv fragments to the minor coat protein p3 of filamentous phage M13 using Sfi I (Nco I)/Not I restriction sites. An amber-stop codon between the scFv-gene and the bacteriophage gene 3 allows the expression of soluble fragment or phage-displayed scFv, in an *E.coli* non-suppressor or suppressor strain, respectively.

Chain shuffling of murine Ki-4 V-genes

The murine Ki-4 scFv was synthesized as described (Klimka et al, 1999). From this scFv, the CDR3-linker-VL-gene fragment was amplified by polymerase chain reaction (PCR) using 30 cycles of 94°C for 1 min, 55°C for 1 min and 72°C for 2 min, with the primers VH-FR3-BACK (5′-GAC ACG GCY GTR TAT TAC TGT-3′) and FD-TET-SEQ (5′-TTT GTC GTC TTT CCA GAC GTT AGT-3′) and the proof-reading *Pfu*-polymerase (Stratagene, La Jolla, CA, USA) according to the manufacturer's instructions. Simultaneously, human VH genes lacking the CDR3-FR4 sequence were amplified from the pHEN1-human scFv repertoire made by Marks et al (1991), using the primers pUC-REV (5′-CAG GAA ACA GCT ATG AC-3′) and VH-FR3-FOR (5′-ACA GTA ATA YAC RGC CGT GTC-3′). For PCR assembly of the amplified fragments, 250 ng of each were combined in a 50 µl mixture and cycled seven times (94°C for 1.5 min, 65°C for 1 min and 72°C for 2 min) to join the fragments. The reaction mixture was then amplified for 30 cycles (94°C for 1 min, 55°C for 2 min and 72°C for 2 min) after the addition of the outer PCR primers pUC-REV/FD-TET-SEQ. Assembly products were digested with Sfi I/Not I and ligated into the phagemid vector pCANTAB6. The ligation mix was purified by phenol extraction and ethanol precipitation and dissolved in 20 µl H_2O. The DNA solution was transfected into 100 µl *E.coli* TG1 by electroporation as described elsewhere (Dower et al, 1988). The cells were grown for 1 h in 2×TY medium at 37°C before plating on 2×TY agar medium containing 100 µg ampicillin ml^{-1} and 2% (w/v) glucose (2×TY-Amp-Glu).

Five different selected human VH genes, determined by DNA-fingerprint analysis as described elsewhere (Marks et al, 1991) from 20 CD30-reactive half-human scFvs, were amplified with primers pUC-REV/VH1-FOR-Xho (5′-CCG CCT CCA CCA CTC GAG ACG GTG ACC GTG GTC CC-3′) using Pfu-polymerase, ligated into pCANTAB6 using restriction enzymes Sfi I/Xho I and electroporated into *E.coli* XL1-Blue. After sequencing of the human VH genes, they were cloned into the pHEN1-VL repertoire (Marks et al, 1991) using the restriction sites Sfi I/Xho I and transfected into *E.coli* TG1 by electroporation as described above. After selection, the human anti-CD30 scFv (hAK30) was finally cloned into pCANTAB6 using the restriction enzymes SfiI/NotI for expression as His-tagged protein.

Selection of phage on the Hodgkin-derived cell line L540

The resulting repertoires of transformed bacteria containing the murine Ki-4 VL linked to human VH repertoire in phagemid vector pCANTAB6 or selected human VH-genes linked to the human VL repertoire in pHEN1, were rescued with helper phage M13K07 as described (Marks et al, 1991). The selection procedure is described elsewhere (Klimka et al, 1999). Briefly, 5×10^5 L540 cells were incubated with 1 ml of 1×10^{13} cfu ml^{-1} phage in 2% (w/v) MPBS (2% Marvel skimmed milk powder in PBS) for 1 h at room temperature (RT). After washing the cells ten times with

© 2000 Cancer Research Campaign

5 ml 2% MPBS and two times with 5 ml PBS by spinning (300 *g*, 3 min, RT) and resuspending respectively, binding phage were eluted with 50 mM HCl and remaining cell debris was spun down (300 *g*, 5 min, RT) after neutralization with 1 M Tris-Cl, pH 7.4. Phage-containing supernatant (SN) was mixed with 3 ml 2×TY-Glu medium and used to transfect logarithmically growing *E.coli* TG1 cells for 30 min at 37°C before plating on 2×TY-Amp-Glu agar medium.

FACS analysis

Cell binding of phage-displayed scFvs was demonstrated by FACS analysis. 5×10^5 L540 target cells were washed in PBS containing 2% (w/v) skimmed milk powder and 0.05% (w/v) sodium azide (2% MPBS/N_3–) and then incubated for 1 h at 4°C with the respective phage or moabs Ki-3 or Ki-4 in 2% MPBS/N_3– respectively. Bound phage were detected with a sheep anti-fd serum (Pharmacia, Uppsala, Sweden; 0.02% (v/v) in 2% MPBS/N_3–) and FITC-labelled rabbit-anti-sheep IgG (Dianova, Hamburg, Germany; 2% (v/v) in MPBS/N_3–). Bound monoclonal antibodies were detected with FITC-conjugated goat-anti-mouse IgG (Becton & Dickinson, Heidelberg, Germany); cells were analysed on a FACScan (Becton & Dickinson). For competition FACS analysis, approximately 10^{12} cfu of phage displaying scFv were mixed with 50 µl of unpurified supernatant from hybridomas secreting moab Ki-3 or moab Ki-4, respectively, resulting in a phage vs moab ratio of approximately 1/1. The mixtures were incubated with the target cells and bound phage were subsequently detected as described.

Sequencing

The scFv-genes were sequenced by the dideoxy chain termination method (Sanger et al, 1977) using Dye-Terminator mix (Perkin Elmer, Norwalk, CO, USA) and the oligonucleotides FD-TET-SEQ and pUC-REV. Products of the sequencing reaction were analysed on a semi-automated ABI Prism sequencer (Perkin Elmer). The nucleic acid sequences of the V regions were compared to the Kabat database of V genes (Kabat and Wu, 1991) and Sanger Centre database (http://www.sanger.ac.uk) to determine the V-gene family and germline V-gene segments.

Purification of recombinant, human, soluble CD30-His

Cloning of the extracellular part of human CD30 receptor fused to a His_6-tag into the eukaryotic expression vector pcDNA3 (Invitrogen, Groningen, The Netherlands) is described elsewhere (Barth et al, 2000). 250 ml supernatant of COS-1 cells transfected with sCD30-His-pcDNA3 plasmid was collected, filtered and incubated with 2 ml Talon™ resin (Clontech, Heidelberg, Germany) for 2 h at 4°C for IMAC purification. The resin was subsequently washed with Tris-buffer (20 mM Trisbase, 100 mM NaCl, pH 8.0) and Tris-buffer, 5 mM Imidazol until the $OD_{280\,nm}$ dropped to 0.001. The sCD30-His protein was then eluted with 250 mM Imidazol in Tris-buffer and dialysed against PBS ovn at 4°C.

Purification of scFv

E.coli HB2151, harbouring the respective scFv genes in pCANTAB6 were used to inoculate 750 ml of 2×TY medium containing 100 µg ml^{-1} ampicillin and 0.1% (w/v) glucose. The culture was grown at 37°C to an $OD_{600\,nm}$ of 0.9 and then supplemented with 1 mM isopropylthio-β-D-galactoside (IPTG) for induction of soluble scFv expression. After 4 h of induction at 30°C, the cells were pelleted and resuspended in 8 ml ice-cold TES (200 mMTris-HCl; 0.5 mM EDTA; 500 mM sucrose). After incubation for 5 min on ice, 8.8 ml of TES/H_2O (1:3) were added and the bacterial suspension was incubated on ice for an additional 20 min. Bacteria were pelleted, SN was collected and the pellet was resuspended in 10 ml TES/15 mM $MgSO_4$ and incubated on ice for 15 min. After centrifugation for 5 min at 300 *g* the supernatants were mixed and centrifuged at 13 000 *g* for 10 min to remove cell debris. The resulting periplasmic fraction was dialysed against Tris-buffer (20 mM Trisbase, 100 mM NaCl, pH 8.0) ovn at 4°C and the scFvs were purified by IMAC using Talon™ resin (Clontech, Palo Alto, USA) as described for the sCD30-His protein.

The human scFv hAK30 was additionally expressed under high-salt stress induction as described elsewhere (Barth et al, 2000). In brief, 2 L bacterial culture were grown at 28°C in TB-medium containing 0.5 mM $ZnCl_2$ and 0.1 M potassium phosphate buffer, pH 7.5 till OD_{600nm} of 1.6. The culture was supplemented with 0.5 M sorbitol, 0.7 M NaCl, 10 mM betain and after 15 min expression was induced by addition of 1 mM IPTG. After overnight growth, bacteria were centrifuged and the pellet was snap-frozen in liquid nitrogen and resuspended in 75 mM Tris-buffer, pH 8.0 containing 10% glycerol, 300 mM NaCl, 2 mM EDTA, 5 mM DTT and Complete™ protease inhibitor (Boehringer Mannheim, Mannheim, Germany). Proteins were extracted by sonification and centrifugation, desalted by gelchromatography using a desalting column (Pharmacia, Uppsala, Sweden) and scFv was isolated by IMAC using Ni-NTA resin (Qiagen, Hilden, Germany).

Eluted protein was thoroughly dialysed against PBS and visualized by gelfiltration, SDS-PAGE and immunoblotting. The final concentration was determined from a scanned Coomassie-stained SDS-PAGE with BSA-dilutions as standards and performing densitometrical analysis with Multi-Analyst software (Bio-Rad, Munich, Germany).

Determination of relative binding affinities of anti-CD30 antibodies

To determine the relative binding affinities of the anti-CD30 antibodies, purified recombinant sCD30-His (70 nM) was incubated for 1 h at RT in duplicates, with dilution series of the respective purified scFvs, the Ki-4 Fab fragment prepared as described elsewhere (Smith, 1993), or the moab Ki-4, respectively. Unbound sCD30-His antigen was detected in a CD30 (Ki-1 antigen)-ELISA kit (DAKO, Glostrup, Denmark) where the coated anti-CD30 antibody BerH2 binds to the same CD30-epitope as the investigated antibodies. sCD30-His captured by BerH2 was detected by peroxidase-conjugated anti-CD30 antibody Ki-1, which binds to a different epitope. The ELISA was performed according to the manufacturer's instructions and extinction at 450 nm was measured. The antibody concentration at which the OD_{450} dropped to 50% of maximum extinction was taken as the apparent K_d.

Measurement of shed sCD30 receptor

2×10^5 L540 cells were washed three times with 10 ml fresh 10% FCS medium and incubated with 1/10 diluted supernatants of

© 2000 Cancer Research Campaign

Table 1 Selection of half-human (A) and human (B) anti-CD30 phage antibodies on Hodgkin cell line L540.

Phage clones	Input titre (cfu)	Output titre (cfu)	Ratio (output/input)	Frequency of positive clones in whole-cell ELISA*
A				
Before selection	–	–	–	0 of 94 (0%)
1st round of selection	4×10^{13}	2×10^{7}	5×10^{-7}	47 of 94 (50%)
2nd round of selection	2×10^{12}	4×10^{9}	2×10^{-3}	75 of 94 (80%)
B				
Before selection	–	–	–	n.d.
1st round of selection	2×10^{13}	2×10^{6}	1×10^{-7}	n.d.
2nd round of selection	4×10^{13}	1×10^{7}	2.5×10^{-7}	0 of 94 (0%)
3rd round of selection	6×10^{13}	5×10^{8}	8.3×10^{-6}	7 of 93 (8%)

*Clones have been stated as positive if $OD_{490\ nm}$ was three times higher than background; cfu, colony forming unit; n.d., not determined.

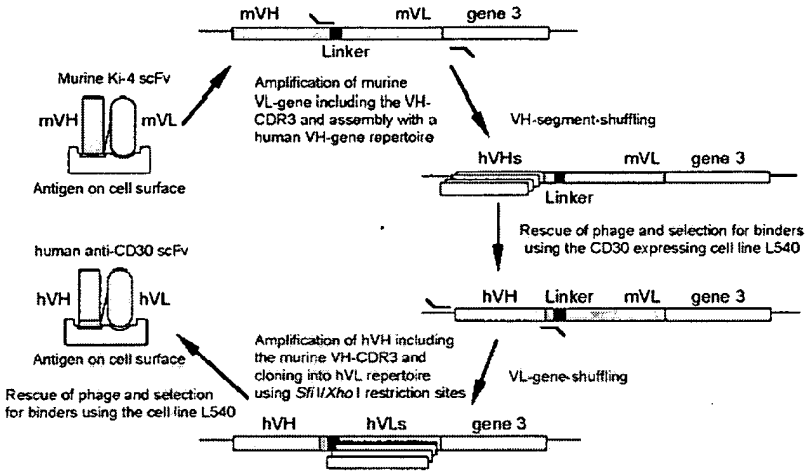


Figure 1 Schematic drawing of the chain-shuffling procedure used for the guided selection of the human anti-CD30 scFv hAK30. The gene 3 encodes the phage minor coat protein p3 and is part of the phagmid vector pCANTAB6.

different anti-CD30 hybridomas or approx. 0.5 µg ml^{-1} purified Ki-4 Fab-fragment, mKi-4 scFv, A12 scFv or hAK30 scFv in 1 ml 10% FCS medium, respectively, to ensure an excess of antibody against the CD30 receptor on the cell surfaces. After 2 h, the cells were washed three times in 10 ml 10% FCS medium by centrifugation (300 *g*) and resuspension to remove unbound antibodies, before the cells were incubated for further 24 h. 100 µl of cell-free supernatants were checked for the level of shed extracellular CD30 receptor using the CD30 (Ki-1 antigen)-ELISA kit (DAKO, Glostrup, Denmark). Relative OD_{450} extinction was determined and compared to the sCD30-level of cells incubated with hybridoma supernatant of an anti-GD2 antibody (BW702) as control.

RESULTS

Cloning of V genes and selection of the half-human and human anti-CD30 scFv

To retrieve a fully human anti-CD30 scFv, the strategy of 'guided selection' (Figure 1) was followed using a recently-cloned murine anti-CD30 scFv (mKi-4 scFv) as guiding molecule. First, the CDR3-linker-VL gene fragment of the murine anti-CD30 Ki-4 scFv was combined with a repertoire of CDR3-truncated human VH genes taken from a repertoire of 1.8×10^{8} human scFv clones (Marks et al, 1991). Phage displaying these combinatorial scFvs were selected for binding to the CD30-positive cell line L540. As documented in Table 1A, two rounds of selection and amplification were sufficient to enrich for CD30-binding, half-human scFv bearing phage (human VH-murine VL) up to 80%, as determined in a whole-cell ELISA using L540 cells. DNA-fingerprint analysis of 12 individual clones with the restriction enzyme BstN I revealed five different patterns in DNA-gel electrophoresis (Figure 2A). The scFv-genes of five representative clones were sequenced and the deduced amino-acid sequences were compared with the Kabat database and the Sanger Centre database of human VH-genes to determine their V-gene family and their closest germline match.

As depicted in Table 2, all five V-genes belong to the VH-I family, with two VH-genes showing the highest homology to the VH DP-75 segment. This segment is also the gene with the highest homology towards the murine VH sequence. The deduced amino-acid sequences of the human and murine Ki-4 heavy-chain CDR1 and CDR2 show a homology of 23–50% (Table 3). However, structural analyses, as far as they can be predicted from the amino-acid sequence (Chothia et al, 1989; 1992), revealed that similar classes of canonical structures for the human and the murine VH-genes occurred.

© 2000 Cancer Research Campaign

Table 2 Deduced amino acid sequences of selected VH- and VL-genes

Heavy chains:

```
                  1        10        20        30        40        50        60        70        80        90        100       110
                  |        | FR1     |         | CDR1    | FR2     | CDR2    |         | FR3     |         | CDR3    |         | FR4
mVH Ki-4          QVKLQESGTELAKPGAAVKMSCKASGYTFT DYWMH WVKQRPGQGLEWIG YINPNTAYTDYNQKFKD KATLTADKSSSTAYMQLRSLTSEDSAVYYCAK KTTQTTWGFPF WGQGTTVTVSS
hVH A9            --Q-----A-VK----S--L---------- EMFI- --R-A-------M- T---SSGS-T-A-R-QG RV-M-R-T-TR-V--E-SR--HD-T------- ---------- ----------
hVH A3            --Q--Q--A-VK----S--V---------- S-GIN -LR-A-------M- G-I-IFGTAN-T-R-Q- RL-I---D-T--S--E-SD-----T------- ---------- ----------
hVH A4            --Q-VQ--A-VK---SS--V------G--S SSSIS --R-A-------M- G-I-SFGTAN-A-R-QG RV-I---E-T-----E-S--R---T-----T- ---------- ----------
hVH A2            --Q-VQ--A-VK--R-S--V---T------ G-Y-- --R-A----F--M- W-D--GGA-T-A---QG RL---R-T-IN----D-SR---D-T------- ---------- ----------
hVH A12(hAK30)    --Q--Q--A-VK---SS--V---T------ G-Y-- --R-A----F--M- W-D--SGA-T-A---QG RLI-SR-T-IN----E--R---D-T------- ---------- ----------
                                      [ H1 ]                    [ H2 ]
```

Light chains:

```
                  1        10        20        30        40        50        60        70        80        90        100
                  |        | FR1     |         | CDR1    | FR2     | CDR2    |         | FR3     |         | CDR3    | FR4
mVL Ki-4          DIVLTQSPKSMAMSVGERVTLSC KASENVDSFVS WYQQKPGQSPKLLIY GASNRYT GVPDRFAGSGSGRDFTLTISSVQAEDLADYHC GQNYRYPLT FGAGTKLEIK
hVL hAK30hVH      ---M----STLSA---D---IT- R--QG-YQMLA --------KA-N---- K--HL-N ---S--S-----T---------L-PD-F-T-Y- Q-LNS---- --G---V---
                                      [ L1 ]                    [ L2 ]          [ L3 ]
```

Amino acid sequences of selected VH and VL fragments are aligned towards parental murine Ki-4 fragments. Different amino acids are indicated. Numbering is according to Kabat and Wu (1997). Definition of CDR-loops (H1, H2, L1, L2, L3) is according to Chothia et al (1989; 1992); FR,. framework region; CDR, complementarity determining region.

© 2000 Cancer Research Campaign

Table 3 V-gene classifications, structural predictions of the CDR-loops and homologies of V-genes involved in the chain-shuffling procedure

V-genes	V-gene family[a]	Predicted canonical structure of CDR loops[b] H/L 1	H/L 2	H/L 3	Human germline gene with closest deduced protein sequence[c]	Amino-acid sequence homology (%) of CDRs towards murine Ki-4
mVH-Ki-4	Mo-VH VII	1	2	n.a.	VH DP-75	100
hVH A9	Hu-VH1	1	2/3	n.a.	VH hv1f10t	36
hVH A3	Hu-VH1	1	2	n.a.	VH VHGL-1.8	32
hVH A4	Hu-VH1	1	2	n.a.	VH DP-10	23
hVH E2	Hu-VH1	1	2/3	n.a.	VH DP-75	50
hVH A12 (hAK30)	Hu-VH1	1	2/3	n.a.	VH DP-75	50
mVL-Ki-4	Mo-VκXXI	2	1	1	Vκ DPK-24	100
hVL-4 (hAK30)	Hu-Vκ1	2	1	1	VκL12a+	41

[a]V-gene families assigned to Kabat database (http://immuno.bme.nwu.edu/famgroup.html); [b]Canonical structures were determined according to Chothia et al (1989; 1992); [c]germline genes assigned to Sanger Centre database (http://www.sanger.ac.uk); n.a., not applicable

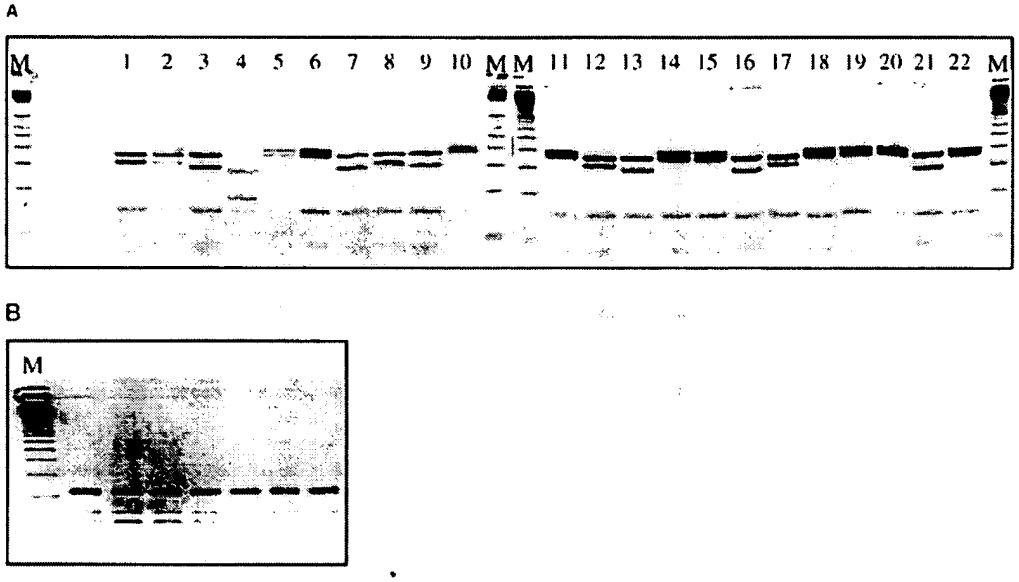


Figure 2 Bst NI fingerprint-analysis of positive scFv clones determined by whole-cell ELISA using cell line L540. The scFv inserts were PCR-amplified from individual colonies using vector-based primers according to Marks et al (1991). The products were digested with Bst NI and analysed on agarose gels. M, 100 bp molecular weight marker. (A) Digests from colonies with half-human scFvs after 1 round of selection (lanes 1 to 10) and 2 rounds of selection (lanes 11 to 22). (B) Lanes 1 to 7 are digest, from colonies with human scFvs after 3 rounds of selection.

The five selected human VH-genes were PCR-amplified and cloned into the phagemid vector pCANTAB6. After sequencing the VH-genes once more, they were pooled and cloned into a 4.5×10^6 member human (h)VL phage antibody library in pHEN-1 (Marks et al, 1991) resulting in a combinatorial library of 8×10^6 individual clones. Three rounds of phage selection and amplification were performed using the Hodgkin-derived cell line L540, which resulted in 8% binders in a whole-cell ELISA (Table 1B). DNA-fingerprint analysis showed that all positive clones contain the same human VH- (from half-human clone A12) and human VL- gene (Figure 2B), which was subsequently confirmed by sequencing two of these clones. The deduced amino-acid sequence of the selected human VL-gene (Table 2) shows that it possesses a 41%-homology to the parental murine VL-gene in the CDR regions and retains similar structural elements (Table 3). The DNA-sequence of the final human anti-CD30 scFv hAK30 was submitted to GenBank (accession number AF117956).

Binding properties of the half-human and human anti-CD30 scFvs

To verify binding specificity of the scFvs against the CD30-epitope recognized by the monoclonal Ki-4 antibody, competition experiments were performed and evaluated by FACS analysis. As shown in Figure 3, binding of the selected scFvs displayed on phage was partially but specifically blocked by the parental Ki-4 moab, but not by the monoclonal Ki-3 antibody, which recognizes a different epitope on the CD30 antigen (shown for mKi-4 scFv, h/mA12 scFv, and human hAK30 scFv). The scF-v-genes were subsequently expressed in *E. coli* non-suppressor strain HB2151

© 2000 Cancer Research Campaign

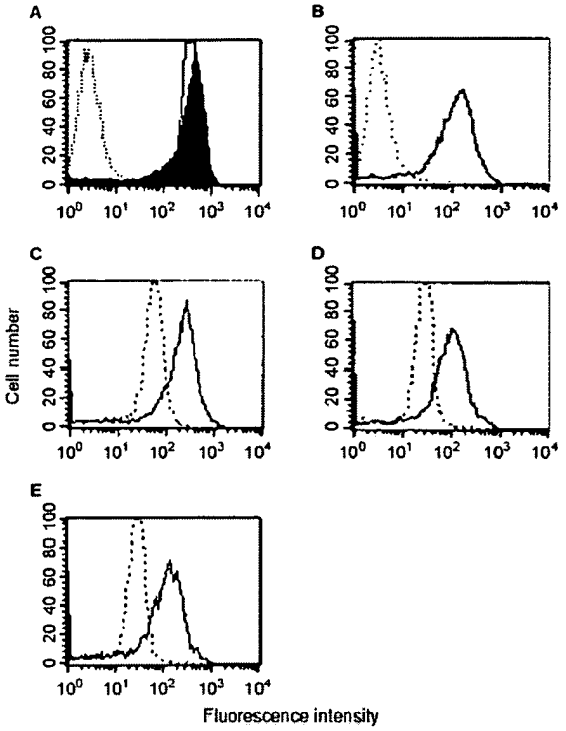


Figure 3 Histograms of FACS-analysis for determination of the CD30-epitope specificity of the selected anti-CD30 scFvs. (A) L540-cells were incubated with anti-CD30 Ki-3- (filled area), Ki-4 (unfilled area) hybridoma-supernatant or PBS (dotted line) and binding was detected by FITC-conjugated goat-anti-mouse IgG antibody. (B) L540 cells were incubated with phage displaying no scFv (dotted line) or phage displaying half-human scFv A12 (black line). L540 cells were incubated with phage-antibodies mKi-4 (C), h/mA12 (D) or hAK30 (E) and additionally with anti-CD30 moab Ki-3- (black line) or Ki-4- (dotted line) hybridoma-supernatant, respectively. Binding of phage-antibodies was subsequently detected using sheep-anti-M13-serum and FITC-conjugated rabbit-anti-sheep-IgG antibody.

and purified by IMAC. The typical yield of purified scFv was approximately 150 µg l^{-1} bacterial culture performing a standard periplasmic extraction, or twice as much using a modified protocol (see Material and methods).

The relative binding affinities were determined by ELISA using a defined concentration of purified sCD30-His protein as antigen and dilution series of the indicated anti-CD30 antibodies (Figure 4). Purified scFv antibodies consisted of at least 95% monomeric molecules demonstrated by gelfiltration (data not shown). Antigen and antibodies were incubated in solution and unbound antigen was subsequently quantified with a coated anti-CD30 antibody, recognizing the same epitope as the investigated antibodies. The antibody concentration at which 50% of the antigen was bound at equilibrium was taken as the apparent K_d. The relative affinities of the moab Ki-4 Fab-fragment, the murine Ki-4 scFv and the half-human anti-CD30 scFv A12 are approximately 10-fold higher than the affinity of the human scFv hAK30 (Table 4), but 10-fold lower than the whole, bivalent monoclonal antibody Ki-4.

Shedding-inhibition of the extracellular part of the CD30 receptor

To investigate the influence of our recombinant anti-CD30 antibodies on the shedding of the extracellular part of the CD30 receptor, L540 cells were incubated with supernatants of different anti-CD30 hybridomas or purified recombinant anti-CD30 scFvs, respectively. After 2 h, the cells were thoroughly washed with

Table 4 Apparent affinities of anti-CD30 antibodies

Anti-CD30 antibody	K_d (M)
moab Ki-4	4×10^{-10}
Ki-4 Fab-fragment	5×10^{-9}
mKi-4 scFv	3×10^{-9}
h/mA12 scFv	7×10^{-9}
hAK30 scFv	3×10^{-8}

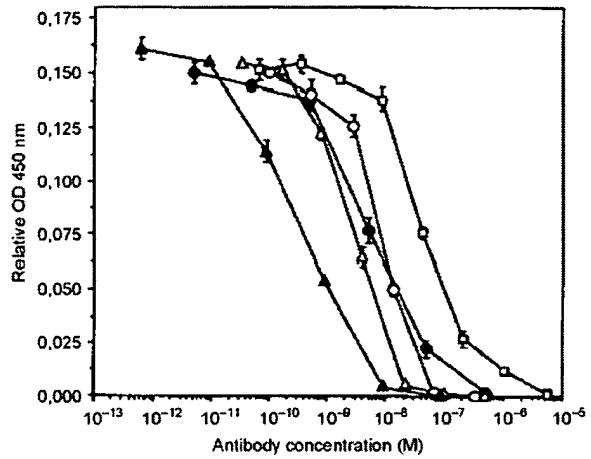


Figure 4 Apparent binding affinities of anti-CD30 antibodies using recombinant sCD30-His antigen. Dilution series of anti-CD30 antibodies moab Ki-4 —▲—, Ki-4 Fab-fragment —●—, mKi-4 scFv —△—, h/mA12 scFv —○— and hAK30 scFv —□—were incubated with recombinant sCD30-His protein and unbound antigen was detected by CD30-ELISA (DAKO).

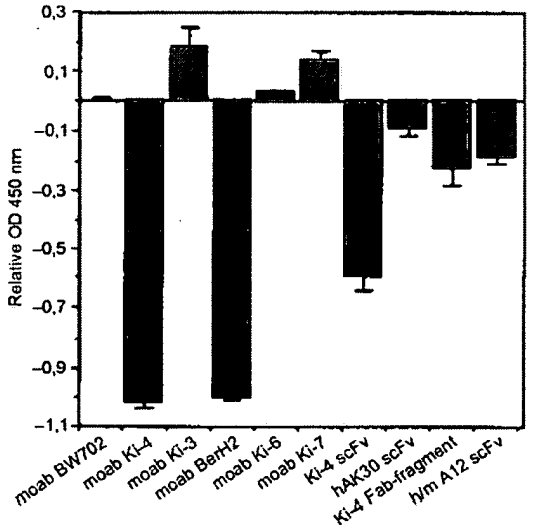


Figure 5 Influence of anti-CD30 antibodies towards naturally occuring CD30 receptor cleavage from Hodgkin-derived L540 cell line. Cells were incubated for 2 h with indicated antibodies and shed CD30 receptor was detected by CD30-ELISA (DAKO). Value of relative OD_{450} of the control, using an irrelevant anti-GD2 antibody (BW702), was set as zero baseline.

© 2000 Cancer Research Campaign

medium to deplete unbound antibodies. After another 24 h of incubation, supernatants were checked for sCD30-level in a CD30 (Ki-1 Antigen) ELISA kit (DAKO) in duplicates. Figure 5 shows that, as was described by Horn-Lohrens et al (1995), Ki-4 and BerH2 strongly inhibit the shedding of the extracellular part of CD30 receptor (sCD30), whereas Ki-3, Ki-6 and Ki-7 increased the sCD30 level to different extents. The recombinant Ki-4-derived anti-CD30 antibodies and the monovalent moab Ki-4 Fab fragment exhibit a comparable inhibition of the cleavage of CD30 receptor from L540 cells, although they are not as potent as the bivalent moabs Ki-4 and BerH2.

DISCUSSION

In this paper, we report the cloning of a human anti-CD30 scFv by guided selection from the murine anti-CD30 scFv Ki-4 using the phage display technology. In contrast to other described procedures (Jespers et al, 1994; Figini et al, 1994), we retained the VH-CDR3 region of the parental murine scFv in this guided selection. This particular region is not only known for its significant importance in determining the binding specificity of an antibody, it is also a highly variable region in every antibody, which makes it less likely to be a major immunogenic part of the molecule. We believe that this region was important in retaining the CD30-epitope specificity of the parental antibody in the human scFv. Watzka et al (1998) describe the humanization through chain-shuffling of an anti-human interferon γ receptor 1 antibody without retaining the VH-CDR3 region of the murine antibody. The resulting fully-human Fab-antibody was antigen specific, but differed in epitope specificity from the parental hybridoma, thereby underlining the importance of the VH-CDR3. This importance of the VH-CDR3 for epitope specificity has also recently been reported by Beiboer et al (2000), thereby confirming our findings.

Selection of phage libraries in this study was performed on a Hodgkin-derived cell line which is known for its high surface expression of CD30 receptor (10^6 receptor molecules per cell), rather than by panning on recombinant CD30 antigen. CD30 is part of the TNF receptor family and many ligands and receptors are known to trimerize. It may therefore not be straightforward to retain natural epitopes on recombinant versions of such cell-surface molecules. Therefore, cell panning on CD30-positive cells is a good and valid alternative. Indeed, using recombinant CD30 protein for panning of several human scFv phage repertoires, to date other groups were unsuccessful in retrieving functional anti-CD30 antibodies.

The selection on cells and therefore native CD30 receptor resulted in five different human VH genes with homology in the CDR1 and CDR2 regions between 23–50%, compared with the parental murine Ki-4 heavy chain. This is similar to what was found in the group of Watzka et al (1998) for their selected anti-human interferon γ receptor antibody (45%), and higher than described in the study of Figini et al (1998), in which an anti-ovarian carcinoma Fab fragment was humanized by guided selection. However, more important than sequence homology might be the length of the CDR regions and the canonical structure of the CDR-loops defined by Chothia et al (1989; 1992). This was striking in the case described by Figini et al (1994) retrieving a human anti-phOx Fab fragment by guided selection which shared these structural elements with the parental mouse antibody. In our study, the selected human VH gene with the highest sequence homology towards mKi-4 VH, but probably slightly different canonical folds, is the one of the half-human anti-CD30 scFv A12. This clone was also predominant after the selection (Figure 2A). Additionally, this human VH gene was selected out of five others after shuffling with the human VL-repertoire. The human VL gene of the finally selected human anti-CD30 scFv has the same length and predicted canonical structure of the CDR-regions as the mKi-4 VL gene, and a 41% homology of the deduced amino-acid sequences concerning these regions. This follows the prediction made by Jespers et al (1994) that there may be a strong preference for retaining V-gene segments with identical canonical folds in guided selection procedures.

Expression of the human anti-CD30 scFv hAK30 as soluble fragment revealed a 10-fold lower apparent K_d for the hAK30 scFv compared with the mKi-4 scFv (Table 4). A loss in affinity after a guided selection procedure has also been reported by other groups (Figini et al, 1998). However, the hAK30 scFv reveals an affinity in the nanomolar range and is therefore expected to be adequate for use as targeting moiety in recombinant immunotherapeutics, in particular when re-formatted as bivalent molecule (Tai et al, 1995). The relative affinity of the monovalent Ki-4 Fab-fragment (5×10^{-9} M) is comparable to the value measured for the mKi-4 scFv and underlines the successful cloning of the functional V-genes from the hybridoma Ki-4. The higher affinity of the bivalent moab Ki-4 (3.7×10^{-10} M) most probably is caused by an avidity effect in the assay.

The monoclonal antibody Ki-4, as well as the moab BerH2, significantly inhibit the naturally occurring shedding of the extracellular part of the CD30 receptor, as demonstrated by Horn-Lohrens et al (1995). Since this is a desired property for an anti-CD30 antibody as part of an immunotherapeutic agent, we were especially interested in retaining the epitope-specificity of the moab Ki-4 in our human anti-CD30 antibody. As shown in Figure 3, the epitope-specificity was retained for the murine and the human scFv. Although the binding of phage antibodies was not completely inhibited by moab Ki-4, which might be due to higher avidity effects of phage (displaying up to five scFv molecules on their surface), competition for binding was not observed by addition of moab Ki-3. The anti-sCD30-shedding property was retained as well, although it was significantly weaker for the monovalent anti-CD30 molecules, which correlates with their apparent binding affinities (Figure 5). A bivalent scFv, like a diabody, may even be as potent as the bivalent moab Ki-4 regarding the CD30-shedding inhibition. Whether the human anti-CD30 scFv hAK30 will be as potent as the murine Ki-4 scFv as part of an anti-CD30 immunotherapeutic agent (Klimka et al, 1999) has to be further analysed, e.g. by fusing it to a human-derived toxin gene (Newton et al, 1996) in order to get a fully human, recombinant immunotoxin.

In summary, we have been able to derive a functional human anti-CD30 scFv (hAK30) from the murine anti-CD30 scFv Ki-4 by guided selection using human V-gene repertoires and phage display technology. The hAK30 scFv retains the epitope specificity of its murine counterpart and inhibits the shedding of the CD30 receptor from the cell surface.

ACKNOWLEDGEMENTS

The authors thank Mrs E van der Linden, Mrs P Henderikx, Mr M Rousch, Mrs G Schön and Mrs S Drillich for their excellent technical assistance. This work was supported in part by 'Deutsche

© 2000 Cancer Research Campaign

Forschungsgemeinschaft, SFB 502' and by 'Boehringer Ingelheim Fond'.

REFERENCES

Barth S, Huhn M, Matthey B, Klimka A, Galinski EA and Engert A (2000) Compatible solute-supported periplasmic expression of functional recombinant proteins under stress conditions. *Appl Environ Microbiol* **66**: 1572–1579

Beiboer SH, Reurs A, Roovers RC, Arends JW, Whitelegg NR, Rees AR and Hoogenboom HR (2000) Guided selection of a pan carcinoma specific antibody reveals similar binding characteristics yet structural divergence between the original murine antibody and its human equivalent. *J Mol Biol* **296**: 833–849

Bosslet K, Mennel HD, Rodden F, Bauer BL, Wagner F, Altmannsberger A, Sedlacek HH, and Wiegandt H (1989) Monoclonal antibodies against epitopes on ganglioside GD2 and its lactones. Markers for gliomas and neuroblastomas. *Cancer Immunol Immunother* **29**: 171–178

Brüggemann M and Neuberger MS (1996) Strategies for expressing human antibody repertoires in transgenic mice. *Immunol Today* **17**: 391–397

Chothia C, Lesk AM, Tramontano A, Levitt M, Smith-Gill SJ, Air G, Sheriff S, Padlan EA, Davies D, Tulip WR and et al (1989) Conformations of immunoglobulin hypervariable regions. *Nature* **342**: 877–883

Chothia C, Lesk AM, Gherardi E, Tomlinson IM, Walter G, Marks JD, Llewelyn MB and Winter G (1992) Structural repertoire of the human VH segments. *Mol Biol* **227**: 799–817

Diehl V, Kirchner HH, Schaadt M, Fonatsch C, Stein H, Gerdes J and Boie C (1981) Hodgkin's disease: establishment and characterization of four in vitro cell lines. *J Cancer Res Clin Oncol* **101**: 111–124

Dower WJ, Miller JF and Ragsdale CW (1988) High efficiency transformation of *E. coli* by high voltage electroporation. *Nucleic Acids Res* **16**: 6127–6145

Figini M, Marks JD, Winter G and Griffiths AD (1994) In vitro assembly of repertoires of antibody chains on the surface of phage by renaturation. *J Mol Biol* **239**: 68–78

Figini M, Obici L, Mezzanzanica D, Griffiths A, Colnaghi MI, Winter G and Canevari S (1998) Panning phage antibody libraries on cells: isolation of human Fab fragments against ovarian carcinoma using guided selection. *Cancer Res* **58**: 991–996

Gruss HJ and Dower SK (1995) The TNF ligand superfamily and its relevance for human diseases. *Cytokines Mol Ther* **1**: 75–105

Hoogenboom HR (1997) Designing and optimizing library selection strategies for generating high-affinity antibodies. *Trends Biotechnol* **15**: 62–70

Hoogenboom HR, de Bruine AP, Hufton SE, Hoet RM, Arends JW and Roovers RC (1998) Antibody phage display technology and its applications. *Immunotechnology* **4**: 1–20

Hoogenboom HR, Lutgerink JT, Pelsers MM, Rousch MJ, Coote J, Van Neer N, DeBruine A, Van Nieuwenhoven FA, Glatz JF and Arends JW (1999) Selection-dominant and nonaccessible epitopes on cell-surface receptors revealed by cell-panning with a large phage antibody library. *Eur J Biochem* **260**: 274–284

Horn-Lohrens O, Tiemann M, Lange H, Kobarg J, Hafner M, Hansen H, Sterry W, Parwaresch RM and Lemke H (1995) Shedding of the soluble form of CD30 from the Hodgkin-analogous cell line L540 is strongly inhibited by a new CD30-specific antibody (Ki-4). *Int J Cancer* **60**: 539–544

Huston JS, McCartney J, Tai MS, Mottola-Hartshorn C, Jin D, Warren F, Keck P and Oppermann H (1993) Medical applications of single-chain antibodies. *Int Rev Immunol* **10**: 195–217

Jespers LS, Roberts A, Mahler SM, Winter G and Hoogenboom HR (1994) Guiding the selection of human antibodies from phage display repertoires to a single epitope of an antigen. *BioTechnology* **12**: 899–903

Kabat EA and Wu TT (1991) Identical V region amino acid sequences and segments of sequences in antibodies of different specificities. Relative contributions of VH and VL genes, minigenes, and complementarity-determining regions to binding of antibody-combining sites. *J Immunol* **147**: 1709–1719

Khazaeli MB, Conry RM and LoBuglio AF (1994) Human immune response to monoclonal antibodies. *J Immunother* **15**: 42–52

Klimka A, Barth S, Matthey B, Roovers R, Lemke H, Hansen H, Arends JW, Diehl V, Hoogenboom HR and Engert A (1999) An anti-CD30 single-chain Fv selected by phage display and fused to *Pseudomonas* exotoxin A (Ki-(scFv)-ETA) is a potent immunotoxin against a Hodgkin-derived cell line. *Br J Cancer* **80**: 1214–1222

Köhler G and Milstein C (1975) Continuous cultures of fused cells secreting antibody of predefined specificity. *Nature* **256**: 495–497

Marks JD, Hoogenboom HR, Bonnert TP, McCafferty J, Griffiths AD and Winter G (1991) By-passing immunization. Human antibodies from V-gene libraries displayed on phage. *J Mol Biol* **222**: 581–597

McCafferty J, Fitzgerald KJ, Earnshaw J, Chiswell DJ, Link J, Smith R and Kenten J (1994) Selection and rapid purification of murine antibody fragments that bind a transition-state analog by phage display. *Appl Biochem Biotechnol* **47**: 157–171

McCall AM, Amoroso AR, Sautes C, Marks JD and Weiner LM (1998) Characterization of anti-mouse Fc gamma RII single-chain Fv fragments derived from human phage display libraries. *Immunotechnology* **4**: 71–87

Meredith RF, Khazaeli MB, Grizzle WE, Orr RA, Plott G, Urist MM, Liu T, Russell CD, Wheeler RH, Schlom J and et al (1993) Direct localization comparison of murine and chimeric B72.3 antibodies in patients with colon cancer. *Hum Antibodies Hybridomas* **4**: 190–197

Newton DL, Xue Y, Olson KA, Fett JW and Rybak SM (1996) Angiogenin single-chain immunofusions: influence of peptide linkers and spacers between fusion protein domains. *Biochemistry* **35**: 545–553

Persic L, Horn IR, Rybak S, Cattaneo A, Hoogenboom HR and Bradbury A (1999) Single-chain variable fragments selected on the 57–76 p21Ras neutralising epitope from phage antibody libraries recognise the parental protein. *FEBS Lett* **443**: 112–116

Sanger F, Nicklen S and Coulson AR (1977) DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci USA* **74**: 5463–5467

Schnell R, Linnartz C, Katouzi AA, Schon G, Bohlen H, Horn-Lohrens O, Parwaresch RM, Lange H, Diehl V, Lemke H and et al (1995) Development of new ricin A-chain immunotoxins with potent anti-tumor effects against human Hodgkin cells in vitro and disseminated Hodgkin tumors in SCID mice using high-affinity monoclonal antibodies directed against the CD30 antigen. *Int J Cancer* **63**: 238–244

Schwab U, Stein H, Gerdes J, Lemke H, Kirchner H, Schaadt M and Diehl V (1982) Production of a monoclonal antibody specific for Hodgkin and Sternberg–Reed cells of Hodgkin's disease and a subset of normal lymphoid cells. *Nature* **299**: 65–67

Schwarting R, Gerdes J, Durkop H, Falini B, Pileri S and Stein H (1989) BER-H2: a new anti-Ki-1 (CD30) monoclonal antibody directed at a formol-resistant epitope. *Blood* **74**: 1678–1689

Shawler DL, Bartholomew RM, Smith LM and Dillman RO (1985) Human immune response to multiple injections of murine monoclonal IgG. *J Immunol* **135**: 1530–1535

Smith TJ (1993) Purification of mouse antibodies and Fab fragments. *Methods Cell Biol* **37**: 75–93

Tai MS, McCartney JE, Adams GP, Jin D, Hudziak RM, Oppermann H, Laminet AA, Bookman MA, Wolf EJ and Liu S (1995) Targeting c-erbB-2 expressing tumors using single-chain Fv monomers and dimers. *Cancer Res* **55** (**23 Suppl**): 5983s–5989s

Watzka H, Pfizenmaier K and Moosmayer D (1998) Guided selection of antibody fragments specific for human interferon gamma receptor I from a human VH- and VL-gene repertoire. *Immunotechnology* **3**: 279–291

Winter G, Griffiths AD, Hawkins RE and Hoogenboom HR (1994) Making antibodies by phage display technology. *Annu Rev Immunol* **12**: 433–455

© 2000 Cancer Research Campaign

doi:10.1006/ mbi.2000.3512 available online at http://www.idealibrary.com on IDEAL® *J. Mol. Biol.* (2000) **2** , 833–849

JMB

Guided Selection of a Pan Carcinoma Specific Antibody Reveals Similar Binding Characteristics yet Structural Divergence Between the Original Murine Antibody and its uman Equivalent

Sigrid . W. Beiboer[1], Anneke Reurs[1], Rob C. Roovers[1] an-Willem Arends[1,2], Nick R. Whitelegg[3], Anthony R. Rees[3] and ennie R. oogenboom[1]*

κ

κ

©

©

Introduction

α

Results

umanisation of the light chain

κ λ

κ ×

λ

×

κ

κ

λ

κ

λ

λ

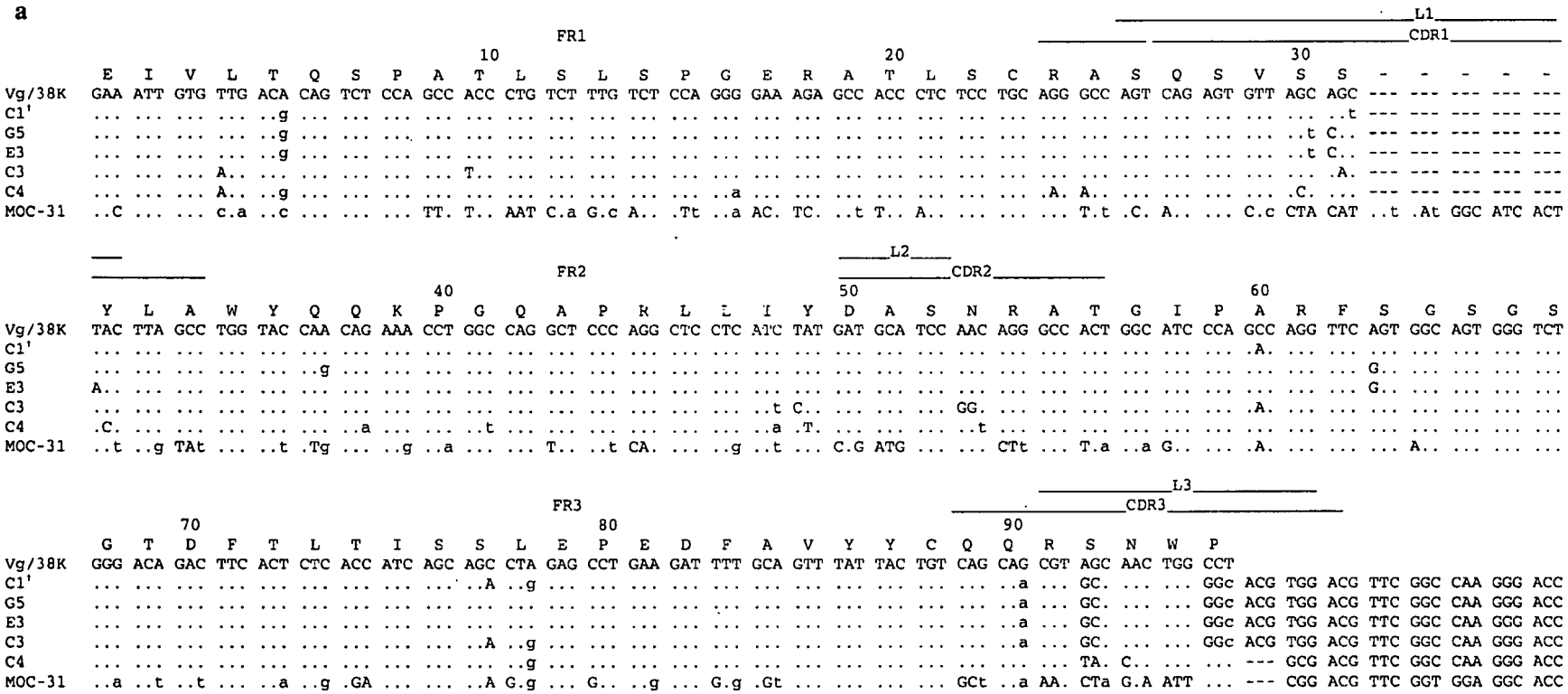
umanisation of the heavy chain

κ

×

κ

a


FR1
L1
CDR1
10
20
30
E I V L T Q S P A T L S L S P G E R A T L S C R A S Q S V S S - - - - -
Vg/38K GAA ATT GTG TTG ACA CAG TCT CCA GCC ACC CTG TCT TTG TCT CCA GGG GAA AGA GCC ACC CTC TCC TGC AGG GCC AGT CAG AGT GTT AGC AGC --- --- --- --- ---
C1' ... ...g ... ...t --- --- --- --- ---
G5 ... ...g ... ...t C.. --- --- --- --- ---
E3 ... ...g ... ...t C.. --- --- --- --- ---
C3 ... ... A.. ... ... T.. ... ... .A. --- --- --- --- ---
C4 ... ... A.. ..g ... ...a ... ... .A. A.. ... ... .C. ... --- --- --- --- ---
MOC-31 ..C ... ... c.a ..c ... ... TT. T.. AAT C.a G.c A.. .Tt ..a AC. TC. ..t T.. A.. ... ... T.t .C. A.. ... C.c CTA CAT ..t .At GGC ATC ACT
L2
CDR2
FR2
40
50
60
Y L A W Y Q Q K P G Q A P R L L I Y D A S N R A T G I P A R F S G S G S
Vg/38K TAC TTA GCC TGG TAC CAA CAG AAA CCT GGC CAG GCT CCC AGG CTC CTC ATC TAT GAT GCA TCC AAC AGG GCC ACT GGC ATC CCA GCC AGG TTC AGT GGC AGT GGG TCT
C1' ... ... .A. ... ...
G5 ... ...g ... ... G.. ... ...
E3 A.. ... ... G.. ... ...
C3 ... ...t C.. ... ... GG. ... ... .A. ... ...
C4 .C. ... ...a ... .t ... ...a .T. ... ...t ... ...
MOC-31 ..t ..g TAt ... ..t .Tg ... ..g ..a ... ... T.. ..t CA. ... ..g ..t ... C.G ATG ... ... CTt ... T.a ..a G.. ... .A. ... ... A.. ... ...
L3
CDR3
FR3
70
80
90
G T D F T L T I S S L E P E D F A V Y Y C Q Q R S N W P
Vg/38K GGG ACA GAC TTC ACT CTC ACC ATC AGC AGC CTA GAG CCT GAA GAT TTT GCA GTT TAT TAC TGT CAG CAG CGT AGC AAC TGG CCT
C1' ... ...A ..g ... ...a ... GC. ... ... GGc ACG TGG ACG TTC GGC CAA GGG ACC
G5 ... ...a ... GC. ... ... GGc ACG TGG ACG TTC GGC CAA GGG ACC
E3 ... ...a ... GC. ... ... GGc ACG TGG ACG TTC GGC CAA GGG ACC
C3 ... ...A ..g ... ...a ... GC. ... ... GGc ACG TGG ACG TTC GGC CAA GGG ACC
C4 ... ...g ... ... TA. C.. ... ... --- GCG ACG TTC GGC CAA GGG ACC
MOC-31 ..a ..t ..t ... ..a ..g .GA ... .A G.g ... G.. ..g ... G.g .Gt ... ... GCt ..a AA. CTa G.A ATT ... --- CGG ACG TTC GGT GGA GGC ACC

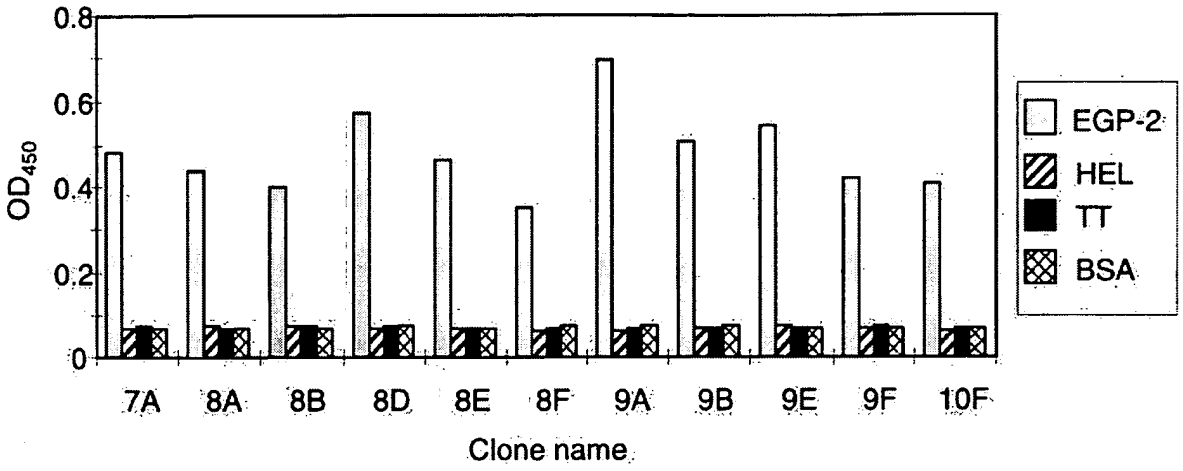
b

	FR1	CDR1	FR2	CDR2
	10 20	30	40	50
Vg/38K	EIVLTQSPATLSLSPGERATLSC	RASQSVSS-----YLA	WYQQKPGQAPRLLIY	DASNRAT
C1'	.....	.....----...	.....	.....
G5	.....	.....R-----...	.....	.....
E3	.....	.....R-----N..	.....	.....
C3	...M.....S.....	.....N-----...	.....H	...G...
C4	...M.....	KT....T.-----S..	.....F	.....
MOC-31	D.....FSNPVTL.TS.SI..	.STK.LLHSNGIT..Y	..L.....S.Q....	QM..L.S

	FR3	CDR3	FR4
	60 70 80	90	
Vg/38K	GIPARFSGSGSGTDFTLTISSLEPEDFAVYYC	QQRSNWP	
C1'	...D.....R.....	...A..GTWT	FGQGTKVEIKR
G5	.....G.....	...A..GTWT	FGQGTKVEIKR
E3	.....G.....	...A..GTWT	FGQGTKVEIKR
C3	...D.....R.....	...A..GTWT	FGQGTKVEIKR
C4	.....	...YH..-AT	FGQGTKVEIKR
MOC-31	.V.D...S.....R..RV.A..VG....	A.NLEI.-RT	FGGGTKLEIKR

†

Discussion


0.8
0.6
0.4
0.2
0
OD450
EGP-2
HEL
TT
BSA
7A
8A
8B
8D
8E
8F
9A
9B
9E
9F
10F
Clone name

μ

κ

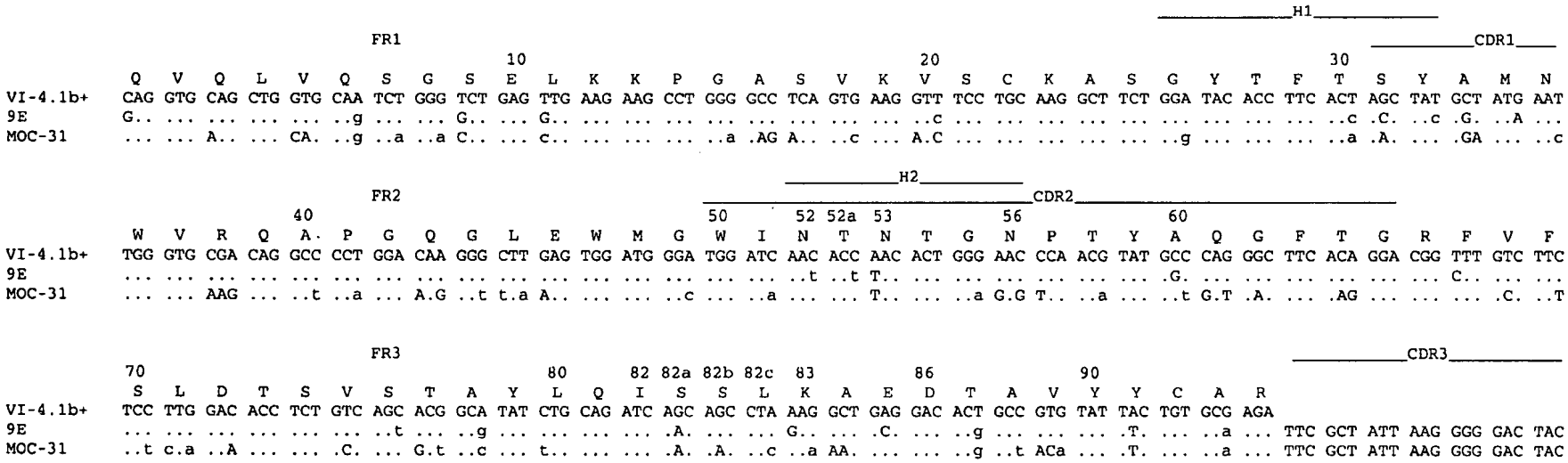
κ

μ

κ
κ
κ
κ
κ
κ

~

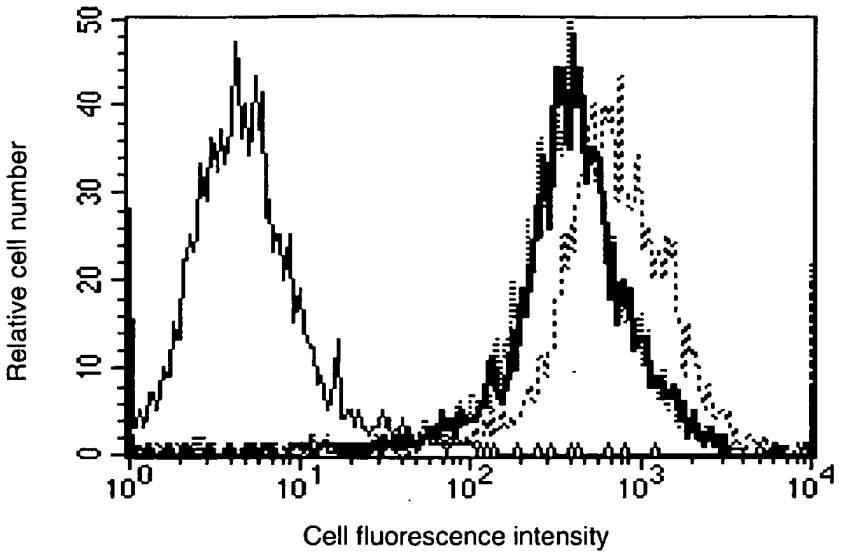
a



b

	FR1 (10–30)	CDR1	FR2 (40)	CDR2 (50 52a 56 60)
VI4.1b	QVQLVQSGSELKKPGASVKVSCKASGYTFT	SYAMN	WVRQAPGQGLEWMG	WINTNTGNPTYAQGFTG
9E	E.......A.V..................	T.GI.	.....	....Y......G.....
MOC-31	..K.Q...P......ET..I.....	N.G..	..K....K..K...	....Y..ES...DD.K.

	FR3 (70–90)	CDR3	FR4
VI4.1b	RFVFSLDTSVSTAYLQISSLKAEDTAVYYCAR		
9E	.L.............N..E.A.....F...	FAIKGDY	WGQGTLVTVSS
MOC-31	..A...E..A.A.....NN..N....T.F...	FAIKGDY	WGQGTTVTVSS



α α

α β

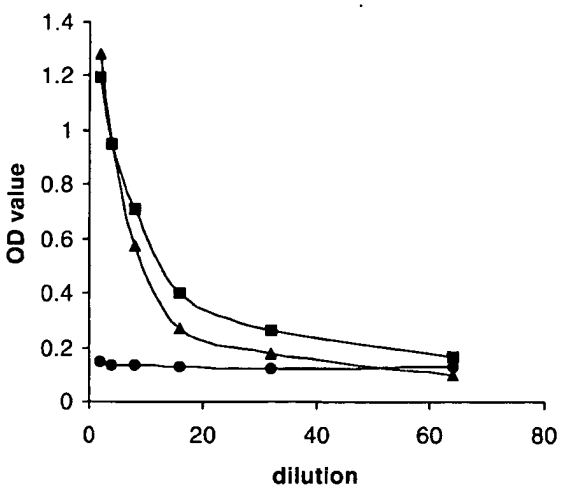
α

α

Materials and Methods

strain

+ +



▲ ■

●

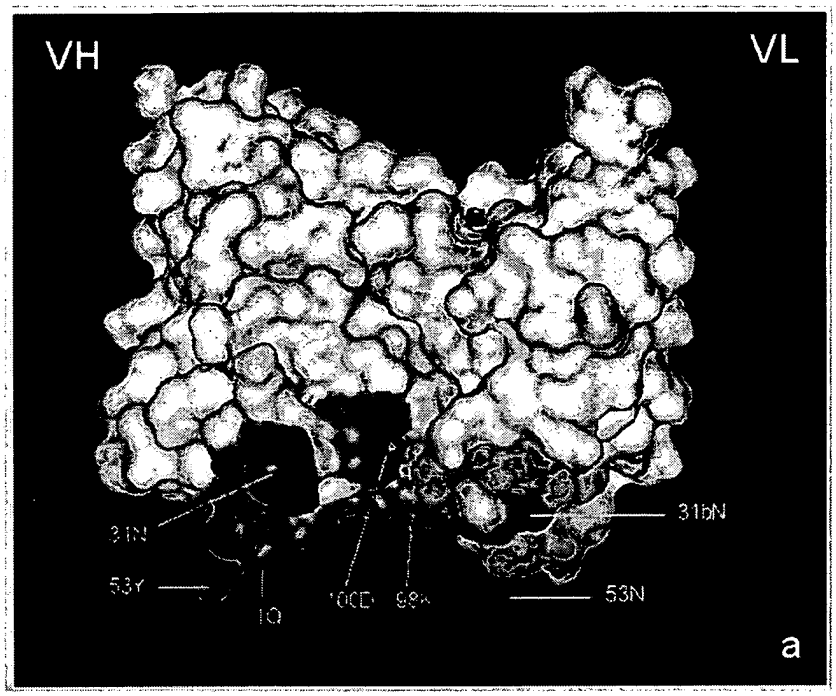
ELISA and kinetic measurement using SPR in BIAcore

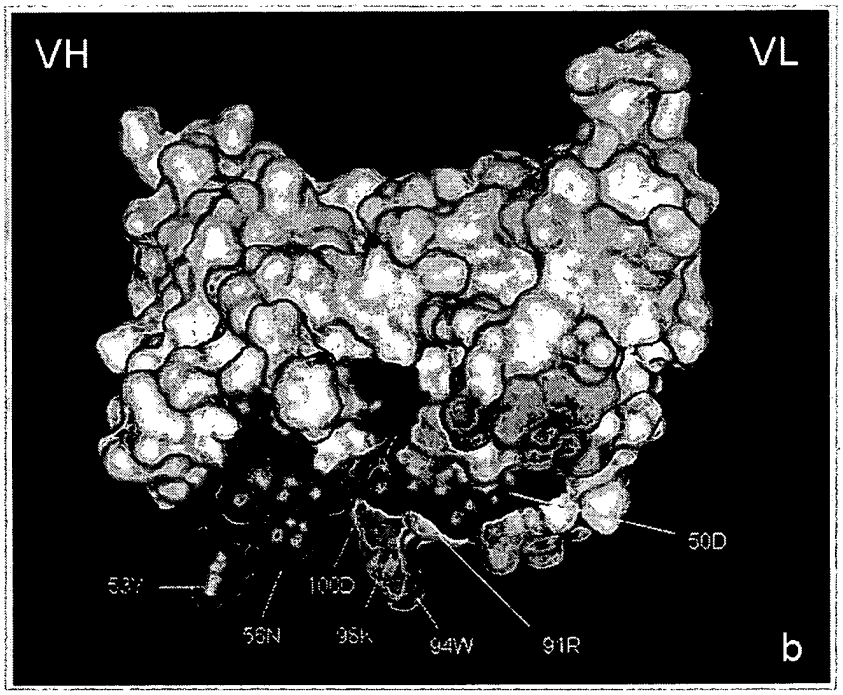
umanisation of light and heavy chains

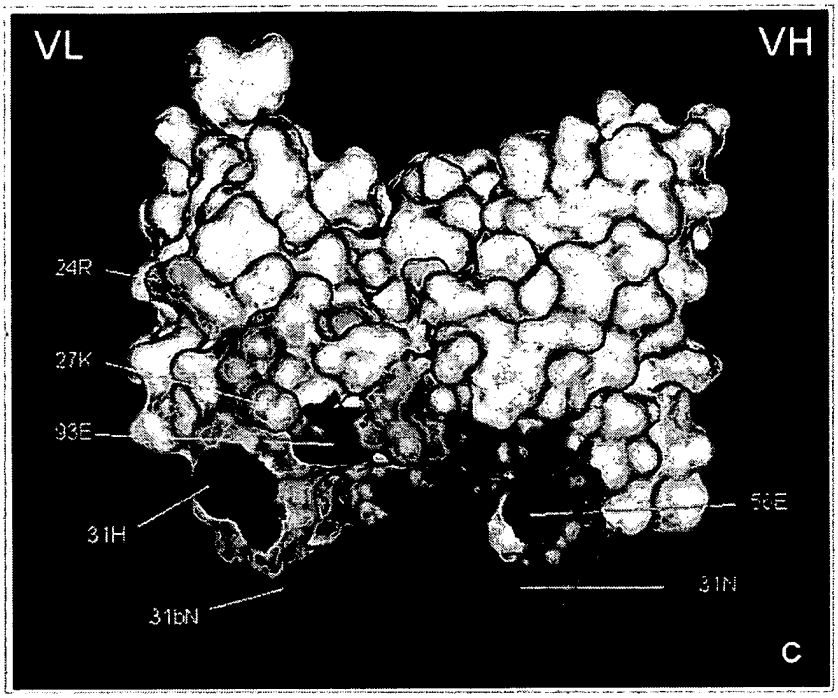
Affinity selection

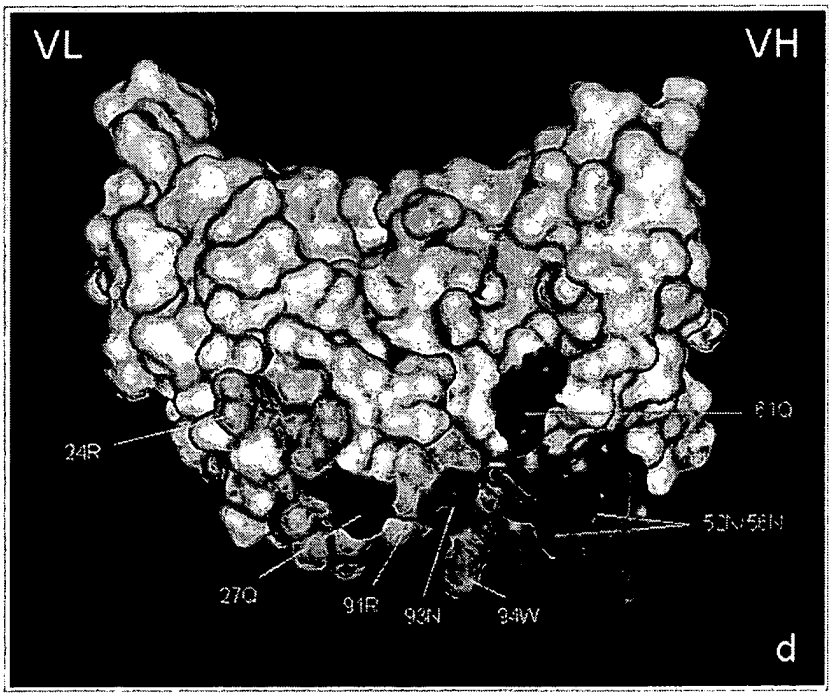
= × +

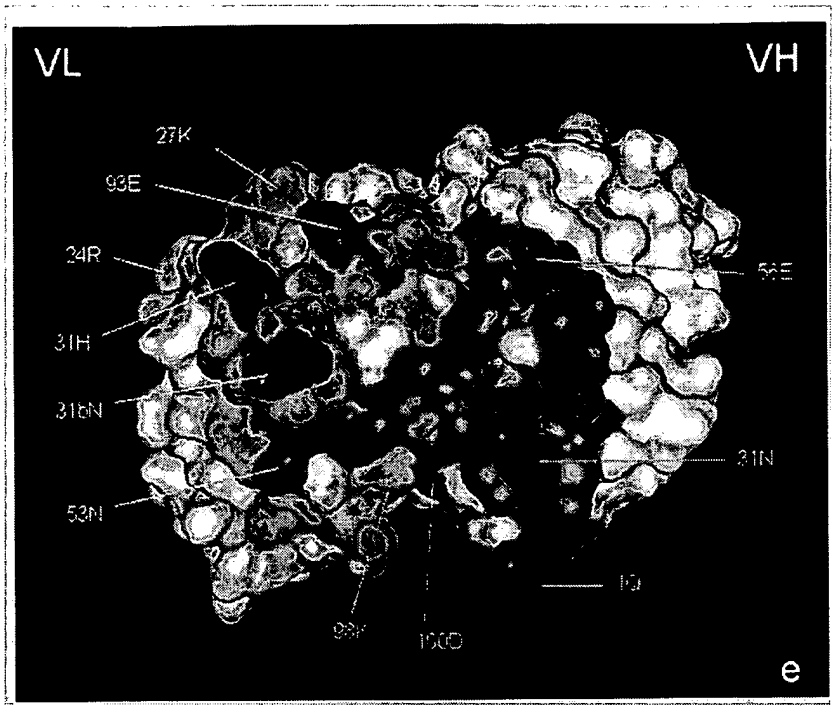
DNA sequence analysis

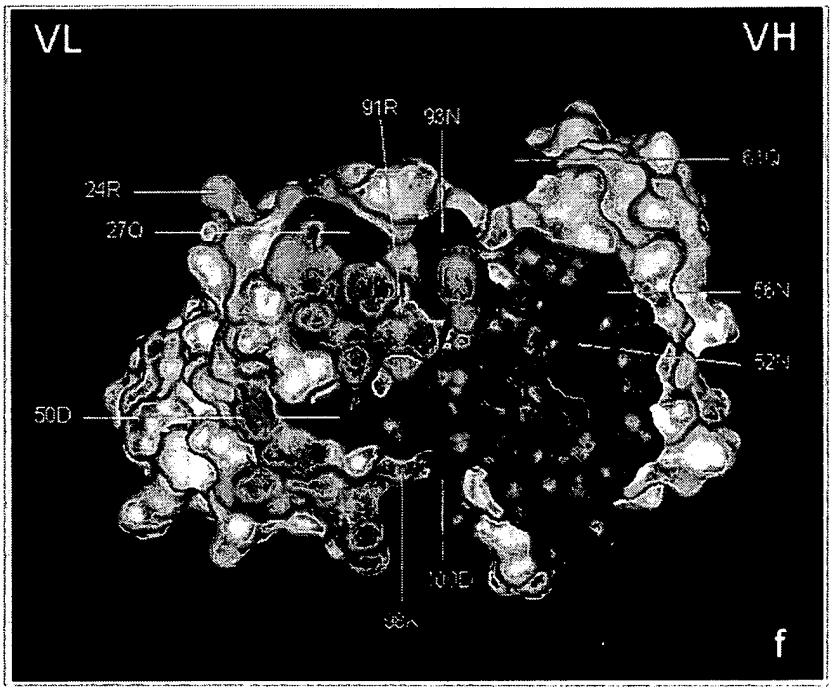

VH
VL
31N
31bN
53Y
53N
a

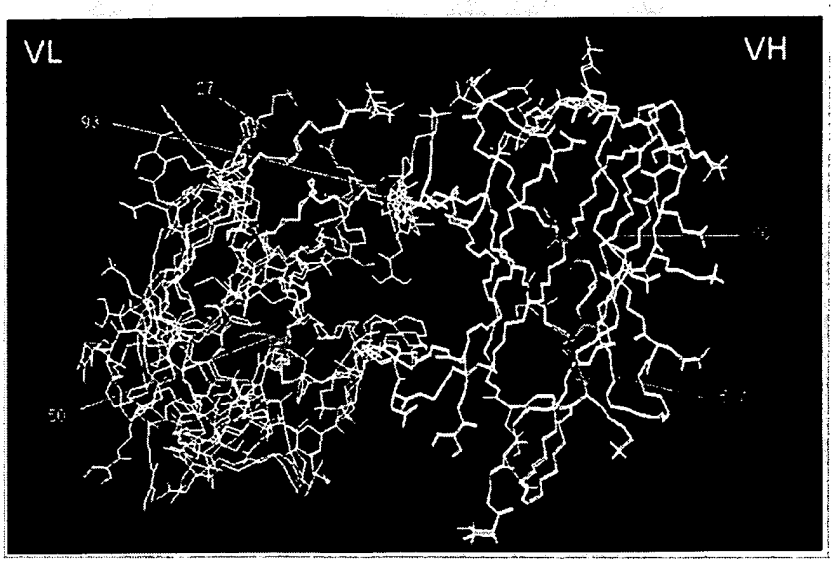

VH
VL
50D
53Y
100D
55N
95K
94W
91R
b


VL
VH
24R
27K
93E
31H
31bN
50E
31N
c


VL
VH
61Q
24R
52N/56N
27Q
91R
93N
94W
d


VL
VH
27K
93E
24R
31H
31bN
53N
56E
31N
e


VL
VH
91R
93N
24R
27Q
50D
52N
f



~ Buil ing loo

Structural modelling

Buil ing th fra or an canonical C

Final cr n

Buil ing non-canonical non- C

$= \Sigma$

References

an non-con r i -chain o lling

FACS analysis and immunohistochemistry

Acknowledgements

Tolerance to Single, but Not Multiple, Amino Acid Replacements in Antibody V_H CDR2

A Means of Minimizing B Cell Wastage from Somatic Hypermutation?[1]

McKay Brown,* Marvin B. Rittenberg,[2]* Ching Chen,[3]* and Victoria A. Roberts[2†]

Mutations in the heavy chain complementarity determining region 2 (CDR2) of the phosphocholine-specific T15 Ab can have a dramatic effect on the ability of the Ab to bind Ag. A panel of multisite mutants that had lost detectable binding to phosphocholine-containing Ags was previously created by saturation mutagenesis of the CDR2 region of T15. Based on the predicted importance of amino acid changes represented in the multisite mutants, we have created single-site mutations, yielding a panel of Abs with which to test 17 of the 19 CDR2 residues. Of the 17 positions examined, only one, Arg^{52}, is intolerant to change, yielding a nonbinder phenotype even with conservative amino acid replacement. Mutation at two other sites, Ala^{50} and Tyr^{55}, can yield a nonbinder phenotype depending on the amino acid replacement. Single-site mutations of the remaining 14 positions allowed retention of binding ability. Thus, except for positions 50, 52, and 55, multiple mutations must be introduced into the CDR2 region to create a nonbinder phenotype. We provide a newly refined model of T15, illustrating the structure and the interactions of the CDR2 region. Our results imply that introduction of point mutations would not normally delete Ag-binding ability until two or more mutations had accumulated. This would minimize potentially harmful effects of somatic mutation on Ig V region genes and improve the chance of survival for an Ab such as T15, which in its unmutated form is already well suited to bind Ag. *The Journal of Immunology,* **1996, 156: 3285–3291.**

The ability of Ab genes to tolerate single-site mutations could provide a mechanism for minimizing B lymphocyte wastage by allowing cells to remain functional while undergoing somatic hypermutation. Supporting this hypothesis, we previously noted that among V_H CDR2 mutants of the T15 anti-phosphocholine (PC)[4] Ab, only 1 of 15 of the Abs with single-site mutations had lost Ag binding ability, whereas 16 of 31 of the Abs with two to four mutations in CDR2 no longer bound Ag (1). In the multisite mutants, it was not clear whether a single critical mutation or a combination of mutations was responsible for the nonbinder phenotype. Here we describe the contribution of single-site mutations in CDR2 to the loss of binding to Ag. The T15 Ab is ideal for these studies, because the heavy chains of anti-PC Abs exhibit little variation in sequence (2), T15 itself does not appear to undergo affinity maturation (3, 4), and T15 provides optimal protection against pathogenic *Streptococcus pneumoniae* (5).

The T15 family of anti-PC Abs has been analyzed extensively. X-ray crystallographic studies of the myeloma protein McPC603, both with and without PC bound in the Ab binding site, revealed residues critical to the interaction with hapten (6–8), and studies of mutant Abs combined with computer modeling of the T15 Ab (9) indicated residues that affect affinity for hapten and carrier determinants (10–13). However, not all of the amino acids forming the active site of the T15 molecule have been examined for their contribution to Ag binding. Such information is crucial for understanding Ag-Ab interactions. We previously examined a library of T15 molecules containing one to four mutations in the V_H CDR2 region (1). As indicated above, over 50% of the mutants either had lost or had reduced capability to bind either PC-protein or the pneumococcus R36a, which expresses PC linked to capsular polysaccharide. These mutagenesis studies of T15 revealed CDR2 residues important to binding and also indicated that highly conserved CDR2 residues could be altered without complete loss of PC binding and without major alteration of Ab structure. Because many of the mutant Abs contained multiple amino acid replacements, the impact of changes at individual positions was deduced by 1) analyzing several Abs with mutations at a given position, and 2) ruling out certain mutations shared by other Abs that retained binding ability. Based on their predicted importance, we selected particular residues for further examination and used site-directed mutagenesis to generate individual replacements. The results corroborate and extend previous findings, and demonstrate the critical contribution of the CDR2 region to the binding site of anti-PC Abs. Furthermore, we found the CDR2 to be relatively insensitive to single substitutions, which has important implications for understanding the potential of somatic mutations to lead to B cell wastage during the immune response.

*Department of Molecular Microbiology and Immunology, Oregon Health Sciences University, Portland, OR 97201; and †Department of Molecular Biology, The Scripps Research Institute, La Jolla, CA 92037

Received for publication November 14, 1995. Accepted for publication February 22, 1996.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

[1] This work was supported by National Institutes of Health Grants AI14985 and AI26827 (to M.B.R.) and GM48877 (to V.A.R.) and National Science Foundation Grant BIR92–23760 (to V.A.R.).

[2] Address correspondence and reprint requests to Dr. Marvin B. Rittenberg, Department of Molecular Microbiology and Immunology, Oregon Health Sciences University, 3181 S.W. Sam Jackson Park Rd., Portland, OR 97201; or to Dr. Victoria A. Roberts, Department of Molecular Biology, The Scripps Research Institute, La Jolla, CA 92037.

[3] Current address: Department of Molecular Biology, Princeton University, Princeton, NJ 08540.

[4] Abbreviations used in this paper: V_H, heavy chain variable region; CDR2, complementarity determining region 2; PC, phosphocholine; Fv, the variable region fragment of an antibody; V_L, light chain variable region.

Copyright © 1996 by The American Association of Immunologists

0022-1767/96/$02.00

Materials and Methods

Site-directed mutagenesis

Mutagenesis was performed in pTZ19 using the Muta-Gene phagemid in vitro mutagenesis kit (Bio-Rad Laboratories, Richmond, CA), with DNA containing the V region of S107 (the V_H gene used by the T15 Ab) as a template and oligonucleotides containing the desired changes. The mutations were verified by dideoxynucleotide chain termination sequencing on double stranded DNA purified from phagemid grown in *Escherichia coli* XL1-blue. T15 wild-type and V_H mutants were cloned into the plasmid pSV2gpt containing the mouse γ2b constant region kindly provided by Dr. J. Sharon. Mutants were resequenced after subcloning into pSV2.

Production of stable transfectants and assay of Ab by ELISA

The heavy chain constructs containing the T15 V_H gene were transfected into the myeloma cell line SP2/0 containing a stably transfected Vκ22Cκ light chain and selected in mycophenolic acid as previously described (1). Transfectants secreting heavy and light chains were analyzed by ELISA for binding to PC-protein and to *S. pneumoniae* strain R36a. Ab was quantified by an ELISA using purified T15 wild-type transfected Ab as a standard as follows. Dilutions of Ab were incubated in ELISA wells coated with purified rabbit anti-mouse IgG. After washing, the wells were incubated with purified rabbit anti-mouse κ coupled to alkaline phosphatase, washed, and read at 410 nm. To determine the ability of the Abs to bind Ag, ELISA plates were coated with PC-histone or *S. pneumoniae* strain R36a (1) and incubated with mutant Ab at 50 ng/ml. Purified alkaline phosphatase-coupled rabbit anti-mouse γ2b was used to detect bound Ab. To determine the percent binding for each mutant Ab, the A410 of mutant Ab at 50 ng/ml was compared with a standard curve of wild-type T15 transfected Ab. Nonbinding to PC-histone was defined as lack of a positive signal from 50 ng/ml of mutant under conditions that routinely allowed positive detection of 0.2 ng/ml (A410>0.05) of wild-type Ab.

Computer modeling of the T15 variable region (Fv)

A previous model of the T15 Fv (9) was constructed with bound sulfate ion. Because many more Ab structures have been determined since that time and because more highly refined coordinates are now available for the McPC603 Ab (kindly provided by Dr. Gerson Cohen, National Institutes of Health), we decided to rebuild the T15 Ab model with bound PC. During construction of the new T15 Ab model, it was superimposed onto a database of crystallographic Ab structures (the ASD). The ASD contains the coordinates of known Ab structures, with the V_L, V_H, and Fv domains separately superimposed (14), allowing comparison of backbone conformations and analysis of the conformational variability of specific side chains. This superposition targeted two regions for refinement, L chain CDR3 and H chain CDR3.

The T15 model was built from the refined McPC603 structure. Side chain replacements for the T15 sequence were made with the graphics program Insight (Biosym Technologies), which builds the new side chain following the dihedral angles of the replaced side chain. The presence of L chain CDR3 residue Pro^{95} in both McPC603 and T15 indicated that the CDR3 loops belong to the same conformational family (15). However, L chain CDR3 residue 90 is Asn in McPC603, but Gln in T15. The terminal amide group of both residues anchors the backbone of CDR3 through a network of hydrogen bonds (15). Analysis of the large number of Ab structures now available, however, indicates that Gln at position 90 causes subtle structural changes compared with Asn, including a shift of the Cα atom (14). Therefore, replacement of the McPC603 Asn^{90} side chain by Gln was followed by adjustment of the side chain dihedral angles. The adjusted conformation was similar to that found in crystallographic structures with Gln^{90} and maintained the hydrogen-bonding network of the terminal amide. The residues on either side of residue 90 are Gln^{89} and Asp^{91} in McPC603, but they are Ala^{89} and Phe^{91} in T15. Building Phe^{91} in the conformation of Asp^{91} in McPC603 resulted in interpenetration of bound PC. The Gln^{89} to Ala change created a hole under the binding site, which was filled by rotation of the Phe^{91} side chain. With this adjustment, the surface of the binding pocket formed by L chain CDR3 was similar to that of McPC603.

In general, H chain CDR3 is the most difficult region to model accurately in the Ab Fv region. In this case, however, nine residues of the 12-amino acid sequence of the T15 CDR3 are identical with those of McPC603, indicating high structural similarity. The three differences are replacement of McPC603 Asn^{95} by Asp, replacement of McPC603 Thr^{100} by Ser, and insertion of Tyr^{100A} in the T15 sequence before Trp^{100B} (indicated in Fig. 1*A* by a W). Examination of the McPC603 structure revealed that residues Gly^{98}, Ser^{99}, and Thr^{100} form a loop at the top of CDR3 with a main chain hydrogen bond between residues Gly^{98} and Thr^{100}. This loop was replaced by a type I β turn made up of residues Gly^{98}, Ser^{99}, Ser^{100}, and inserted residue Tyr^{100A}.

The PC hapten was added to the model by superposition of the main chain framework atoms of the crystallographic structure of McPC603 with bound PC (2 MCP in the Protein Data Bank) (16) onto the T15 model. PC fit well into the T15 model, retaining the van der Waals and hydrogen-bonding contacts observed in McPC603. The system was minimized in a step-wise fashion similar to that of the original T15 model (9) with the program Discover (Biosym Technologies, San Diego, CA). Residues Tyr^{97} to Tyr^{100A} of H chain CDR3 were first relaxed by energy minimization with the rest of the system held fixed. The Ab and hapten were then surrounded by a 5-Å shell of waters, which was relaxed while the protein and hapten were held fixed. Finally, the entire system was relaxed with the protein and PC strongly forced to their original positions using a harmonic potential (force constant of 1000 kcal/Å) during 1000 steps of steepest descent minimization.

The final minimized structure was close to that of the McPC603 crystallographic structure. The root-mean-square deviation of all backbone atoms (N, Cα, C, O) between McPC603 and the T15 model was 0.31 Å (H chain CDR3 residues 97–100A of T15 were excluded from this calculation). The root-mean-square deviation of all nonhydrogen atoms of the starting T15 model (which was identical with McPC603 except where side chains were replaced) to the final, energy-minimized model was 0.36 Å. The hydrogen-bonding network of the Gln^{90} side chain in L chain CDR3 matched that of Asn^{90} in McPC603. In addition, the Gln^{90} Cα atom shifted about 0.5 Å, consistent with the change previously observed in analysis of Ab structures (14).

Results

Experimental strategy

Previous data obtained from T15 Ab mutants generated by saturation mutagenesis of the CDR2 region of the V_H gene implicated several positions as crucial to Ag binding (1). The multisite mutant Abs displaying a nonbinder phenotype indicate side chains that could have a critical function in Ag binding (Table I). To test the relative contribution of each of the CDR2 sites predicted to influence binding and to assess whether the effects of each mutation were additive, we introduced single-site mutations into the germline T15 V_H gene at 10 of the 19 CDR2 residues by oligonucleotide-directed mutagenesis. In some cases, we introduced single-site mutations that were not represented in the multisite mutant panel; these mutations were made either to analyze a particular site in greater detail (i.e., A50E, S, V, and G) or as a result of the use of degenerate oligonucleotide primers in the mutagenesis reaction (i.e., S62C). These mutants plus the single-site mutants previously generated by random mutagenesis (1) yielded a panel of Abs with point mutations in 17 of the 19 CDR2 residues (Table II). Mutant Abs were analyzed by ELISA for binding to PC-histone and to *S. pneumoniae* strain, R36a.

Mutations in the first loop of H chain CDR2, residues 50 to 58

Residues 50 to 58 of the H chain constitute part of the Ag-binding site. Crystallographic structures of Ag-Ab complexes frequently show extensive contacts between Ag and this region of CDR2 (17, 18), indicating its importance in Ag binding. Therefore, we expected that single-site mutation of many of these residues would have a profound effect on Ag binding and would be responsible for the loss of binding observed in some of the multisite mutants obtained by saturation mutagenesis (1).

The double site mutant M22 contains a replacement at H chain Ala^{50} and displays no detectable binding to either PC-protein or pneumococcus R36a (Table I). To test whether the substitution of Ala^{50} alone could cause loss of binding and to investigate the steric and electrostatic requirements of residue 50, five single-site mutants were made (Table II). Replacement of Ala^{50} by the polar side chains Thr, Glu, or Ser gave mutant Abs with no detectable binding to pneumoccus or PC-protein. Replacement by Gly permitted

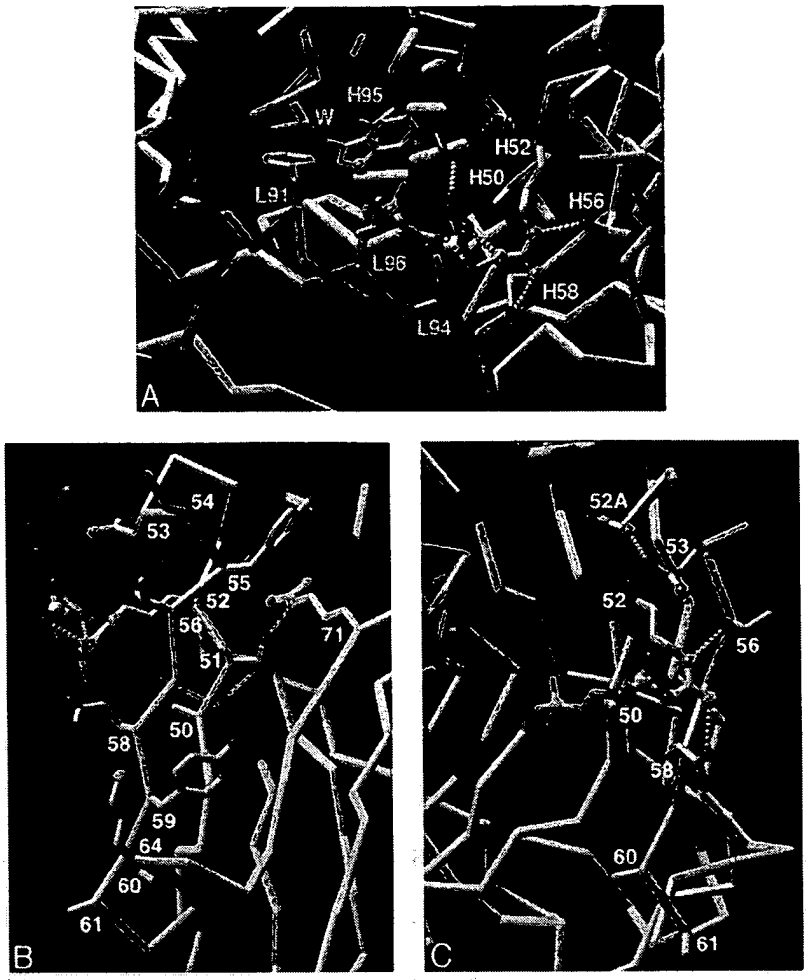


FIGURE 1. Computer model of the T15 Ab showing the structure and interactions of V_H CDR2. The T15 Fv is displayed as a Cα trace with selected side chains. Side chain polar atoms are shown as colored spheres: N is blue; and O is red. The V_L domain (magenta framework residues) and V_H domain (light blue framework residues) have colored CDRs: CDR1 is red, CDR2 is yellow, and CDR3 is green. V_L side chain labels are preceded by L, and V_H side chain labels are preceded by H. V_H CDR3 residue Trp100B is labeled W. PC is shown as a ball and stick model with atoms colored: C is green, N is blue, O is red, and P is yellow. Selected hydrogen bonds are shown as lavender spheres. *A*, V_H residue Ala^{50} is surrounded by critical residues involved in an extensive hydrogen-bonding network in the T15 combining site. This view from solvent looking into the Ag-binding site shows Ala^{50}-contacting residues V_H Arg^{52} (hydrogen bonded to both PC and V_H residues Thr^{56} and Glu^{58}), V_H Glu^{35}, and V_L Tyr^{94}, which are essential for PC binding. *B*, V_H CDR2 residues Ser^{51}, Tyr^{55}, and Tyr^{59} extend away from the T15 Ag-binding site and interact with the V_H framework. The view is rotated approximately 90° from that in *A*, with the binding site at the *top left*, the V_L chain not shown, and the intermolecular salt-bridge between PC (*top left*) and Arg^{52} displayed to orient the viewer. Bound PC is far from CDR2 residues 59 to 65. Also shown is an H bond between Ser^{51} and Arg^{71} of the framework that is believed to be important in stabilizing the CDR2 backbone. *C*, A hydrogen bond between Asn^{52A} and Asn^{53} stabilizes the conformation of the first loop region of V_H CDR2. The view is rotated 90° from that in *B*, with bound PC in the *center*, the V_L domain not shown, and the intramolecular hydrogen bonds of Arg^{52} displayed.

Table I. *Binding of T15 CDR2 multisite mutants*

CDR2	50 A	51 S	52 R	a N	b K	c A	53 N	54 D	55 Y	56 T	57 T	58 E	59 Y	60 S	61 A	62 S	63 V	64 K	65 G	Binding[a] R36a	Binding[a] PC-histone
Wild type	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	100	100
M22	T	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	L	–	–	0	0
M241	–	R	–	–	–	–	–	E	–	–	–	–	S	–	–	–	–	–	–	0	0
M229	–	G	–	–	–	–	–	–	–	–	–	–	–	–	P	–	–	–	–	0	0
M34	–	–	K	–	–	–	–	–	–	–	–	–	–	–	–	P	–	N	–	0	0
M85	–	–	I	–	–	–	–	–	H	–	–	–	–	–	–	–	–	–	–	0	0
M32	–	–	–	K	–	–	–	–	–	–	–	–	–	R	–	–	–	–	–	0	12[b]
M154	–	–	–	K	–	–	–	–	–	I	–	–	–	–	–	–	–	–	–	0	0
M166	–	–	–	K	–	–	–	–	–	–	–	G	–	–	–	Y	M	–	–	0	0
R52K-E58Q	–	–	K	–	–	–	–	–	–	–	–	Q	–	–	–	–	–	–	–	0	0

[a] Numbers indicate percentage of wild-type transfected Ab binding determined by ELISA. Zero equals no detectable binding, which is <0.4% of wild type. The data represent the average of 2 to 3 determinations.
[b] Data from Chen et al., 1992.

retention of only 1% of the wild-type binding for PC-histone. Interestingly, the mutant with Ala^{50} replaced by Val, which is isosteric with Thr, retained 10% of the wild-type binding, indicating that hydrophobic side chains can be tolerated at position 50.

Two multisite mutants, M241 and M229, had replacements at Ser^{51} to Arg and Gly, respectively, and were unable to bind Ag, while the single-site mutation to Ile^{51} retained measurable binding (1). Thus, to examine further the contribution of Ser^{51}, we made single-site mutations of Ser^{51} to Arg and Gly. All single-site mutants retained measurable binding activity (Table II), showing that the loss of binding in the multisite mutants cannot be explained by mutation at position 51 alone.

In the saturation mutagenesis study, all seven multisite mutants containing replacement of Arg^{52} (with Lys, Ile, Ser, or Gly) lacked detectable Ag binding (1). The corresponding single-site mutants as well as a Thr^{52} mutant, all displayed the nonbinder phenotype (Table II). The Arg^{52} side chain is involved in an extensive hydrogen-bonded network (Fig. 1), including a salt bridge with the hapten phosphate group and an intramolecular salt bridge with the Glu^{58} side chain. Because of the reduced hydrogen bonding ability

Table II. *Binding of T15 CDR2 single-site mutants*[a]

	Mutation	R36a	PC-histone
Wild type		100	100
A50	T	0	0
	E	0	0
	S	0	0
	V	10	9
	G	0	1
S51	R	7	16
	G	42	28
	I	14	19[b]
R52	K	0	0
	I	0	0
	S	0	0
	G	0	0
	T	0	0
N52a	K	7	9
	T	20	27
	H	16	20[b]
N53	K	57	67
	S	10	27[b]
	T	7	39[b]
D54	V	77	76[b]
Y55	S	0	0[b]
T56	I	4	9
	R	41	50
	K	12	14
T57	A	114	114[b]
E58	Q	78	126[b]
Y59	S	43	42
S60	I	50	57[b]
A61	P	38	33
	S	62	43
	T	67	50
	G	71	77[b]
S62	Y	53	43
	C	65	57
	P	7	100[b]
V63	F	57	65[b]
	A	52	63[b]
	E	64	73[b]
K64	T	77	47
G65	D	73	58[b]

[a] Numbers indicate percentage of wild type as in Table I. Standard deviations <30% of the value shown
[b] Data from Chen et al., 1992.

of the Lys side chain compared with Arg, Lys^{52} may be unable to satisfy both inter- and intramolecular bonding. Replacement of the Glu^{58} side chain would remove the potential intramolecular salt bridge with the Lys^{52} side chain, possibly freeing Lys^{52} for interaction with the hapten. We tested this hypothesis by replacing Glu^{58} with the noncharged residue Gln, creating the double mutant R52K-E58Q (Table I). However, like the single-site mutants at position 52, the double mutant displayed no detectable binding (Table I), an effect that cannot be ascribed to the single E58Q mutation (Table II).

Multisite mutants with replacements in H-chain CDR2 residues in the loop at the top of the binding site gave variable results (1). The multisite mutants with Lys in position 52A (M154, M166, and M32) showed no detectable binding or weak binding to PC-histone; the one with Thr (M244) had detectable binding to R36a and PC-histone (1). The single-site mutants with Asn^{52A} replaced by Lys, Thr, or His displayed a drop in activity, but still retained some binding ability (Table II), indicating that replacement of residue 52A contributes to the decreased binding observed for the multisite mutants. Replacement of the neighboring residue Asn^{53} by Lys caused less than a twofold decrease in binding compared with wild-type, although replacement by Ser or Thr caused larger decreases (Table II). Because the loss of binding for multisite mutant M154 could not be explained solely by the mutation of residue 52A, we replaced Thr^{56} by Ile, the only other mutation in M154. This mutant had <10% of wild-type binding (Table II). We also replaced Thr^{56} by Arg because MOPC 167, another PC-binding myeloma that uses the same V_H gene, has Arg at position 56. This single-site mutant retained approximately 50% of wild-type binding; however, replacement by Lys, which also has a positively charged side chain, showed only about 10% of wild-type binding (Table II). Thus, the nonbinder phenotype of multisite mutants with mutations at position 52A appears to result from the combination of mutations in at least two sites.

Mutations in the second loop of H chain CDR2, residues 59 to 65

The loop formed by residues 59 to 65 lies outside the combining site. Although there is much sequence variability in this region among Abs belonging to different subgroups, there is very little sequence variation among the V_H gene subgroup IIIA (which includes T15), even though binding specificities differ (2). In the T15 Ab, single substitutions were tolerated, with little to moderate loss of binding function in all positions of this CDR2 loop formed by residues 59 to 65 (Table II). One residue, Tyr^{59}, is conserved (97%) in all V_H regions. Mutation of the invariant Tyr^{59} heavy chain residue to Asp (1) or Ser (Table II) results in only a moderate (60%) decrease in Ag binding, which suggests that conservation of this position among Abs may be due to a function more critical to Ab folding, assembly, secretion, or stability, rather than to Ag binding. As discussed above, the binding loss of double site mutant M229 (Table I) is not solely due to replacement of Ser^{51} by Gly (Table II). The other mutation in M229, Ala^{61} to Pro, could influence Ag binding indirectly by constraining the conformation of the second CDR2 loop. The Pro^{61} single-site mutant displayed about one-third of the wild-type binding (Table II). Replacement of Lys^{64} by Thr also allowed substantial binding (Table II), although Lys^{64} is conserved in CDR2 regions consisting of 19 amino acids.

The T15 Ab model

The computational model of the T15 Fv domain preserves the features found in the McPC603 crystallographic structure. The identical residues surrounding bound PC in McPC603 and in the T15 model, L chain Tyr^{94} and Leu^{96} and H chain Glu^{35}, Ala^{50}, Arg^{52}, Thr^{56}, Glu^{58}, and Trp^{100B}, superimpose closely (root-mean-square deviation of 0.30 Å for all nonhydrogen atoms). The hydrogen-bonding network involving L chain residue Tyr^{94} and H chain residues Glu^{35}, Arg^{52}, Thr^{56}, and Glu^{58} in McPC603 is preserved in the T15 model, including the salt bridge between Arg^{52} and PC (Fig. 1*A*). L chain Asp^{91} in McPC603 is believed to help stabilize the bound positively charged choline group of PC. In T15 (Fig. 1*A*), L chain residue 91 is Phe, but the loss of the Asp^{91} group may be compensated for by the presence of H chain Asp^{95} (19), which is Asn in McPC603. These changes have little effect on the binding site, as evidenced by the small change in PC position (root-mean-square deviation of 0.68 Å for the nonhydrogen atoms of PC calculated after superposition of the conserved main chain atoms of the T15 model and McPC603). The T15 model also preserves the side chain hydrogen-bonding interactions between Ser^{51} and framework residue Arg^{71} (Fig. 1*B*) and between Asn^{52A} and Asn^{53} at the top of the first CDR2 loop (Fig. 1*C*).

Discussion

Structural studies on the PC-binding myelomas McPC603 and T15 indicate that the binding cavities are geometrically and electrochemically suited for interaction with the charged hapten (6, 9, 19).

Although the hapten-binding pocket is formed by the hypervariable loops from both heavy and light chains, the dominant role of the H chain (20, 21) is emphasized by its contribution of contact residues for the phosphate group, which largely determines the orientation of the hapten in the binding pocket (11). The role of the H chain CDR2 region in providing a good fit for PC Ag has been deduced by x-ray crystallographic studies (6, 7), chemical modification of the binding cavity (22), and site-specific mutagenesis studies (11).

In addition to its importance to Ag binding, the CDR2 region is particularly suited for mutation studies (1, 23) because it contains residues that span three distinct structural environments: a structurally conserved region (residues 50–52 and 56–58), a solvent-exposed loop at the top of the binding site likely to interact with carrier protein (residues 52A-55), and a loop distant from the Ag-binding site (residues 59–65). No other CDR spans all three types of regions. The uniqueness of CDR2 is further accentuated by the very different lengths predicted by structural variation (residues 52A-55) (15) and by sequence variation (residues 50–65) (24) for this hypervariable loop. Heavy chain CDR2 plays an important role in forming the Ag-binding pocket and has extensive contacts with Ag in many Ag-Ab complexes (17, 18), including McPC603 (6) and T15 (9).

From the panel of multisite mutants, we found that changes in the CDR2 region of T15 V_H did not affect the overall topology of Ab structure, as indicated by retention of binding by anti-Id Abs. However, mutations in CDR2 can have a dramatic effect on the ability to bind Ag (1). In the present study, the mutation of individual CDR2 residues confirmed the critical function of several residues postulated to be important by analysis of the previous multisite T15 CDR2 mutants. This panel of T15 single-site mutants covers all three regions of CDR2 defined above, clarifying the role of each region of CDR2 in Ag binding.

Mutation of the structurally conserved V_H residues 50 to 52 and 56 to 58 can have a profound effect on Ag binding. The main chain hydrogen bonds between residues 52 and 56 and between residues 50 and 58 form part of the conserved β sheet structure of the V_H domain. The side chains of residues 50, 52, 56, and 58 extend toward the V_L domain, forming part of the Ag-binding region, and the side chains of residues 51 and 57 extend away from V_L, with Ser^{51} making extensive contacts with the V_H framework (Fig. 1*B*). Of these six residues, Ala^{50} and Arg^{52} are most sensitive to mutation, probably because of their position near the center of the binding site. Although Ala^{50} does not contact Ag, its small size may be requisite for shaping the Ag-binding cavity, as indicated by crystallographic studies of the McPC603 Ab (8). Ala^{50} is conserved in all Abs that bind PC. It contacts both the conserved framework and the side chains of L chain residue Tyr^{94} and H chain residues Arg^{52} and Glu^{35} in McPC603 (6) and T15 (Fig. 1*A*). These three residues have been shown to be critical for PC binding (1, 11, 13) and are part of an extensive hydrogen-bonding network at the bottom of the binding pocket (Fig. 1*A*). Ala^{50} can be replaced by Val or by Gly without complete loss of binding, but not by Thr, which is isosteric with Val (Table II). Thus, introduction of a polar atom is detrimental, possibly due to disruption of the surrounding hydrogen bond network.

Arg^{52} has been shown to be essential for Ag binding in anti-PC Abs (11), and this position is invariant in all PC-binding Igs (2). The contribution of Arg^{52} to the binding energy of interaction with PC is greater than that of all of the other residues combined (25). The Arg^{52} side chain forms a salt bridge with the hapten phosphate group and has extensive intramolecular contacts, including a charged hydrogen bond with the Thr^{56} side chain and a salt bridge with the Glu^{58} side chain (Fig. 1*A*). Even the conservative change to a positively charged Lys prevented binding (Table II), corroborating a similar finding for the McPC603 Ab (11). Removal of the intramolecular salt bridge (in the R52K-E58Q double mutant), which may compete with formation of the intermolecular salt bridge, did not restore activity. Thus, the lack of binding of the Lys^{52} mutant may be due to an inability of the shorter Lys side chain to satisfy the stringent directional and spatial requirements for salt bridge interaction with hapten.

The side chain of residue Ser^{51} extends away from the Ag-binding site, lying between the backbone of heavy chain CDR2 and the framework. It contacts the side chains of H chain residues Tyr^{55} and Arg^{71} (Fig. 1*B*), which are important for maintaining the backbone structure of CDR2 (15, 26). This positioning and the sequential placement of Ser^{51} between Ala^{50} and Arg^{52} suggest that mutation of Ser^{51} could have a large effect on Ag affinity. Mutants with replacement of Ser^{51} retained Ag-binding function, however, suggesting that the surrounding, conserved β sheet structure partially insulates the binding pocket from structural perturbations due to mutation of Ser^{51}.

The interaction of residues Thr^{56} and Glu^{58} with Arg^{52} suggests that replacement of these two residues could also have a profound effect on Ag binding. Even though Glu^{58} is invariant in all PC binding proteins (2), replacement by Gln is not detrimental, indicating that its negatively charged side chain does not play a critical role in binding. All three single-site mutants of Thr^{56} show detectable binding, with the replacement by the hydrophobic residue Ile resulting in the poorest binding. Thus, although the bonding of Thr^{56} and Glu^{58} with Arg^{52} may have a stabilizing effect, it does not appear to be critical to Ag binding.

The first loop of the H chain CDR2 region is formed by sequence-variable residues and one residue (Tyr^{55}) that is conserved. Residues Asn^{52A} through Tyr^{55} comprise the six-residue hairpin turn characteristic of type 4 hypervariable heavy chain loops in CDR2 (27). In these structures, the amino acids Gly, Asn, Asp, and Pro play an important role in allowing the hairpin turn conformation (26). In McPC603 and T15, a hydrogen bond between the side chains of Asn^{52A} and Asn^{53} (Fig. 1*C*) may assist in turn formation. The putative role of this solvent-exposed loop in recognizing carrier determinants (28) is supported by the differential binding of mutants N53S and N53T to PC-protein compared with their binding to the pneumococcal strain R36a. In addition, it has been shown that mutation of Asn^{52A} to His in the T15 V_H gene converts an Ab from a nonbinder to one that strongly binds the PC epitope in the bacterium *Proteus morganii* (29). However, none of the Abs in our T15 mutant panel, including those substituted at position 52A, react with a *P. morganii* extract (kindly provided by L. Claflin) in an ELISA (data not shown), possibly due to the different L chains used. The position of residue 53 near the apex of the Ag-binding pocket may be particularly important in distinguishing carrier determinants in conjunction with the PC-hapten. Asp^{54}, which follows the β hairpin, appears tolerant to change; replacement of Asp^{54} by Val has little effect on binding. In addition, the PC-binding myeloma protein McPC603 has a positively charged Lys at this position (2). This tolerance to mutation is consistent with the placement of Asp^{54} on the edge of the binding pocket (Fig. 1*B*). However, mutation of Asp^{54} caused improved binding to PC-protein in a related Ab (30), suggesting that this residue can have a significant influence on binding PC Ags. The conserved amino acid Tyr^{55} is thought to be crucial for maintenance of the canonical CDR2 structure (15). This residue is conserved in Abs with CDR2 regions that are 19 amino acids in length, such as T15 (2). Conservative replacements permit retention of binding (1), but replacement by Ser results in a nonbinder phenotype (Table II). The importance of Tyr^{55} in maintaining the structural geometry of

the CDR2 backbone has been illustrated in the computer model of T15 (1).

Residues 59 to 65 form a second loop region in CDR2 that is distant from the Ag binding site (Fig. 1, *B* and *C*). One of these residues, Tyr^{59}, is highly conserved among all Igs, but the rest of the loop can be highly variable in sequence. It is unclear whether there is a structural or a functional basis for the sequence variability of residues 60 to 65 (2). In the T15 Ab, these residues do not show a dramatic effect on binding when mutated individually, but when combined with substitutions at positions outside this region, they may abolish binding completely, as in M229. Previous studies have suggested that residues 60 to 65 do not form Ag contacts, but a recent report showed that substitution of this loop in a humanized murine Ab resulted in improved Ag binding (31). Although the authors suggest that this may be due to improved contacts between residues 60 to 65 and the Ag, the crystallographic studies of the humanized Ab (in the absence of Ag) show a shift of 1.5 Å in the position of heavy chain framework residue 47 as a result of substituting the human residues (31). This suggests that residues 60 to 65 may affect Ag binding through long range conformational effects, as proposed by Chen et al. (30).

Analysis of the contributions of individual sites to the nonbinder phenotype reveals that binding loss can occasionally be attributed to one critical site, as is the case with T15 mutants containing changes in positions Ala^{50} or Arg^{52}. The binding loss in other mutants can be explained by the additive effects of individual mutations, as in mutant M154. In some cases, however, the decreased binding of multisite mutants is considerably greater than the individual mutations would predict, as would be expected if the amino acids are close to each other or can act in concert to create a local deleterious structural change (32). For example, even though residues 51 and 59 do not contact each other in the wild-type Ab (Fig. 1*B*), they both lie between the framework residues and the backbone of the structurally conserved region of CDR2 composed of residues 50 to 52 and 56 to 58. Thus, the substitutions at residues 51, 54, and 59 in M241 may effect a concerted structural change that could be beyond what the conserved backbone structure can tolerate. Other mutants, such as M229, have mutations that are distant from each other (positions 51 and 61; Fig. 1*B*) but together cause loss of binding, implying that long range interactions are responsible for the loss of binding.

Studies of mutants combined with computer modeling of the combining site can help elucidate the nature of the interaction of each residue with Ag and its influence on surrounding structure. With this interdisciplinary approach, structural requirements of residues that do not contact Ag but form a support for the combining site can be determined. Understanding the role of such residues is essential in attempts to redesign Abs for increased affinity or to replace CDR regions successfully with those of other species. Conversely, this type of analysis may define areas that can accommodate changes in sequence without detrimental effects on binding.

Results from several studies suggest that the T15 Ab does not undergo affinity maturation (3, 4, 33). Analysis of randomly generated mutants in the T15 V_H CDR2 demonstrated that 43% of the Abs lost Ag-binding ability, suggesting that mutations in this region may contribute to B cell wastage during the immune response (1). However, with one exception, the dramatic loss of binding function was restricted to Abs containing more than one mutation. It was not clear whether the low frequency of detrimental mutation in the randomly generated single mutants (1 of 15) was due to lack of substitution at critical positions. By mutating the positions known to affect binding dramatically in the multiple mutants, we show that, in general, the CDR2 region is tolerant to single mutations; thus the nonbinder phenotype usually must result from the cumulative effect of mutations at more than one site. Undetectable binding could be ascribed to the substitution of a single residue at only 2 of the 10 positions that putatively cause the nonbinder phenotype represented by the multiple mutants listed in Table I. Even markedly different replacements at most positions (such as introduction of an aromatic side chain, a proline, or a change in charge) did not result in complete loss of binding.

It has been shown that B cells can proliferate substantially in germinal centers before the onset of somatic mutation (34). Thus, expansion of cells expressing unmutated V genes coupled with the likelihood that single amino acid changes are not detrimental to B cell survival would provide a larger B cell pool for the selective process. These results imply that B cell wastage resulting from somatic mutation during an immune response would be low until introduction of a second substitution. Although in this study, we analyzed only the CDR2 region of the T15 H chain, we have obtained similar results by mutation of the framework 2 region of the T15 H chain as well as the CDR2 region of an unrelated Ab PCG1–1 (G. D. Wiens, et al., in preparation). Evolutionarily, this mutational tolerance would be a way of minimizing loss of Ab-producing cells in vivo during B cell differentiation, providing a second chance for the mutational mechanism to produce a more structurally favorable Ab.

Acknowledgments

We are grateful to Tom Macke for help with the computer graphics, to Dr. Gerson Cohen for providing updated coordinates of the McPC603 Ab, to Dr. Jacqueline Sharon for the pSV2gpt vector, and to Dr. Latham Claflin for the *P. morganii* extract. We also thank Drs. Greg Wiens, Elizabeth Whitcomb, Mary Stenzel-Poore, Tammy Martin, and Kurt Heldwein for review of the manuscript.

References

1. Chen, C., V. A. Roberts, and M. B. Rittenberg. 1992. Generation and analysis of random point mutations in an antibody CDR2 sequence: many mutated antibodies lose their ability to bind antigen. *J. Exp. Med. 176:855.*
2. Kabat, E. A., T. T. Wu, H. M. Perry, K. S. Gottesman, and C. Foeller. 1991. *Sequences of Proteins of Immunological Interest*, 5th Ed. U.S. Department of Health and Human Services, Bethesda.
3. Claflin, J. L., and J. Berry. 1988. Genetics of the phosphocholine-specific antibody response to *Streptococcus pneumoniae*: germline but not mutated T15 antibodies are dominantly selected. *J. Immunol. 141:4012.*
4. Malipiero, U. V., N. S. Levy, and P. J. Gearhart. 1987. Somatic mutation in anti-phosphorylcholine antibodies. *Immunol. Rev. 96:59.*
5. Briles, D. E., C. Forman, S. Hudak, and J. L. Claflin. 1982. Anti-phosphorylcholine antibodies of the T15 idiotype are optimally protective against *Streptococcus pneumoniae. J. Exp. Med. 156:1177.*
6. Satow, Y., G. H. Cohen, E. A. Padlan, and D. R. Davies. 1986. Phosphocholine binding immunoglobulin Fab McPC603: an x-ray diffraction study at 2.7A. *J. Mol. Biol. 190:593.*
7. Segal, D. M., E. A. Padlan, G. H. Cohen, S. Rudikoff, M. Potter, and D. R. Davies. 1974. The three dimensional structure of a phosphorylcholine-binding mouse immunoglobulin Fab and the nature of the antigen binding site. *Proc. Natl. Acad. Sci. USA 71:4298.*
8. Padlan, E. A., D. R. Davies, S. Rudikoff, and M. Potter. 1976. Structural basis for the specificity of phosphorylcholine-binding immunoglobulins. *Immunochemistry 13:945.*
9. Chien, N. C., V. A. Roberts, A. M. Giusti, M. D. Scharff, and E. D. Getzoff. 1989. Significant structural and functional change of an antigen-binding site by a distant amino acid substitution: proposal of a structural mechanism. *Proc. Natl. Acad. Sci. USA 86:5532.*
10. Diamond, B., and M. D. Scharff. 1984. Somatic mutation of the T15 heavy chain gives rise to an antibody with autoantibody specificity. *Proc. Natl. Acad. Sci. USA 81:5841.*
11. Glockshuber, R., J. Stadlmuller, and A. Pluckthun. 1991. Mapping and modification of an antibody hapten binding site: a site-directed mutagenesis study of McPC603. *Biochemistry 30:3049.*
12. Kobrin, B. J., S. Buhl, M. J. Shulman, and M. D. Scharff. 1991. A V region mutation in a phosphocholine-binding monoclonal antibody results in loss of antigen binding. *J. Immunol. 146:2017.*
13. Rudikoff, S., A. Giusti, W. D. Cook, and M. D. Scharff. 1982. Single amino acid substitution altering antigen binding specificity. *Proc. Natl. Acad. Sci. USA 79: 1979.*

14. Roberts, V. A., J. Stewart, S. J. Benkovic, and E. D. Getzoff. 1994. Catalytic antibody model and mutagenesis implicate arginine in transition-state stabilization. *J. Mol. Biol. 235:1098.*
15. Chothia, C., and A. M. Lesk. 1987. Canonical structures for the hypervariable regions of immunoglobulins. *J. Mol. Biol. 196:901.*
16. Bernstein, F. C., T. F. Koetzle, G. J. B. Williams, J. E. Meyer, M. D. Brice, J. R. Rodgers, O. Kennard, T. Shimanouchi, and M. Tasumi. 1977. The Protein Data Bank: a computer-based archival file for macromolecular structures. *J. Mol. Biol. 112:535.*
17. Davies, D. R., E. A. Padlan, and S. Sheriff. 1990. Antibody-antigen complexes. *Annu. Rev. Biochem. 59:439.*
18. Wilson, A., and R. L. Stanfield. 1993. Antibody-antigen interactions. *Curr. Opin. Struct. Biol. 3:113.*
19. Padlan, E. A., G. H. Cohen, and D. R. Davies. 1985. On the specificity of antibody/antigen interactions: phosphocholine binding to McPC603 and the correlation of three dimensional structure and sequence data. *Ann. Inst. Pasteur Immunol. 136C:271.*
20. Barstad, P., S. Rudikoff, M. Potter, M. Cohn, W. Konigsberg, and L. Hood. 1974. Immunoglobulin structure: amino terminal sequences of mouse myeloma proteins that bind phosphorylcholine. *Science 183:962.*
21. Goetze, A. M., and J. H. Richards. 1977. Structure-function relationships in phosphorylcholine-binding mouse myeloma proteins. *Proc. Natl. Acad. Sci. USA 74: 2109.*
22. Grossberg, A. L., L. M. Krausz, L. Rendina, and D. Pressman. 1974. The presence of arginyl residues and carboxylate groups in the phosphorylcholine-binding site of mouse myeloma protein, HOPC8. *J. Immunol. 113:1807.*
23. Casson, L. P., and T. Manser. 1995. Random mutagenesis of two complementarity determining region amino acids yields an unexpectedly high frequency of antibodies with increased affinity for both cognate antigen and autoantigen. *J. Exp. Med. 182:743.*
24. Kabat, E. A., T. T. Wu, and H. Bilofsky. 1977. Unusual distribution of amino acids in complementarity-determining (hypervariable) segments of heavy and light chains of immunoglobulins and their possible roles in specificity of antibody combining sites. *J. Biol. Chem. 252:6609.*
25. Novotny, J., R. E. Bruccoleri, and F. A. Saul. 1989. On the attribution of binding energy in antigen-antibody complexes McPC603, D1.3, and HyHEL-5. *Biochemistry 28:4735.*
26. Tramontano, A., C. Chothia, and A. M. Lesk. 1990. Framework residue 71 is a major determinant of the position and conformation of the second hypervariable region in the VH domains of immunoglobulins. *J. Mol. Biol. 215:175.*
27. Chothia, C., A. M. Lesk, A. Tramontano, M. Levitt, S. J. Smith-Gill, G. Air, S. Sheriff, E. A. Padlan, D. Davies, W. R. Tulip, P. M. Colman, S. Spinelli, P. M. Alzari, and R. J. Poljak. 1989. Conformations of immunoglobulin hypervariable regions. *Nature 342:877.*
28. Claflin, J. L., J. Wolfe, A. Maddalena, and S. Hudak. 1984. The murine antibody response to phosphocholine. In *The Biology of Idiotypes.* M. I. Greene and A. Nisonoff, eds. Plenum Press, New York, p. 171.
29. Claflin, J. L., J. George, C. Dell, and J. Berry. 1989. Patterns of mutations and selection in antibodies to the phosphocholine-specific determinant in *Proteus morganii. J. Immunol. 143:3054.*
30. Chen, C., V. A. Roberts, S. Stevens, M. Brown, M. P. Stenzel-Poore, and M. B. Rittenberg. 1995. Enhancement and destruction of antibody function by somatic mutation: unequal occurrence is controlled by V gene combinatorial associations. *EMBO J. 14:2784.*
31. Eigenbrot, C., T. Gonzalez, J. Mayeda, P. Carter, W. Werther, T. Hotaling, J. Fox, and J. Kessler. 1994. X-ray structures of fragments from binding and nonbinding versions of a humanized anti-CD18 antibody: structural indications of the key role of VH residues 59 to 65. *Proteins Struct. Funct. Genet. 18:49.*
32. Wells, J. A. 1990. Additivity of mutational effects in proteins. *Biochemistry 29: 8509.*
33. Rodwell, J. D., P. J. Gearhart, and F. Karush. 1983. Restriction in IgM expression. IV. Affinity analysis of monoclonal anti-phosphocholine antibodies. *J. Immunol. 130:313.*
34. Jacob, J., J. Przylepa, C. Miller, and G. Kelsoe. 1993. In situ studies of the primary immune response to (4-hydroxy-3-nitrophenyl)acetyl. III. The kinetics of V region mutation and selection in germinal center B cells. *J. Exp. Med. 178: 1293.*

ACADEMIC PRESS

Available online at www.sciencedirect.com

SCIENCE DIRECT®

Biochemical and Biophysical Research Communications 307 (2003) 198–205

BBRC

www.elsevier.com/locate/ybbrc

A peptide mimetic of an anti-CD4 monoclonal antibody by rational design

Florence Casset,[a] Florence Roux,[a] Patrick Mouchet,[a] Cedric Bes,[b] Thierry Chardes,[c] Claude Granier,[b] Jean-Claude Mani,[b] Martine Pugnière,[b] Daniel Laune,[b] Bernard Pau,[b] Michel Kaczorek,[a] Roger Lahana,[a] and Anthony Rees[a,*]

[a] *Synt:em, Parc Scientifique Georges Besse, FR-30035 Nîmes Cédex 1, France*

[b] *Centre de Pharmacologie et Biotechnologie pour la Santé, CNRS UMR 5094, Faculté de Pharmacie, 15 Avenue Charles Flahault, 34093 Montpellier Cedex 5, France*

[c] *Laboratoire de Pathologie Comparée, INRA-cNRS UMR 5087, 30380 Saint Christol les Alès, France*

Received 20 May 2003

This article is dedicated to the memory of Jean-Claude Mani, who passed away before this work was completed.

Abstract

The development of rational methods to design 'continuous' sequence mimetics of discontinuous regions of protein sequence has, to now, been only marginally successful. This has been largely due to the difficulty of constraining the recognition elements of a mimetic structure to the relative conformational and spatial orientations present in the parent molecule. Using peptide mapping to determine 'active' antigen recognition residues, molecular modeling, and a molecular dynamics trajectory analysis, we have developed a peptide mimic of an anti-CD4 antibody, containing antigen contact residues from multiple CDRs. The design described is a 27-residue peptide formed by juxtaposition of residues from 5 CDR regions. It displays an affinity for the antigen (CD4) of 0.9 nM, compared to 2 nM for the parent antibody ST40. Nevertheless, the mimetic shows low biological activity in an anti-retroviral assay.

© 2003 Elsevier Science (USA). All rights reserved.

Keywords: Antibody modeling; Antibody mimetic; Paratope mimetic; Rational design; Anti-CD4 monoclonal antibody; ST40

The design of small oligopeptide, or peptide-like mimics to reproduce the activity of large natural proteins has numerous applications in both therapeutics and diagnostics. Where the activity region of interest is located within a continuous sequence of the protein, methods have been described in which the conformational requirements of such regions can be met by various chemical cross-linking or other synthetic strategies [1,2]. However, where the activity region(s) of a protein consist of discontinuous segments of the polypeptide chain, the problem of mimetic design is particularly acute.

Several approaches have been proposed to mimic the discontinuous binding surface of a protein. Mutter has defined a template approach (TASP, for template-assembled synthetic proteins) where amino acids defined as bioactive are linked to a topological template to mimic the protein activity [3,4]. Others have used small multi-disulfide-containing *mini-proteins* (protease inhibitors and animal toxins) as templates to reproduce the binding surface of a particular protein by mutation of amino acids at the surface of the mini-protein. In one exemplification of this approach, functional sites of CD4 were transferred to a scorpion toxin generating an inhibitor of the HIV-1 gp120-CD4 interaction [5]. Non-peptidyl mimics, in which the spatial arrangement of important amino acid side chains of the parent protein are reproduced, has been also described for protein A [6]. A particularly interesting approach, in which a synthetic cyclic mimic of a discontinuous binding site

* Corresponding author. Fax: +33-4-66-04-86-67.
E-mail address: trees@syntem.com (A. Rees).

0006-291X/03/$ - see front matter © 2003 Elsevier Science (USA). All rights reserved.
doi:10.1016/S0006-291X(03)01131-8

from interleukin-10 was generated using peptide libraries in a multi-step sequence, was described by Reineke et al. [7]. In all, 446 combinations of cyclization were synthesized and tested. The final mimic displayed good activity. Using a similar approach, a peptido-mimetic with a high affinity to rheumatoid arthritis-associated Class II major histocompatibility (MHC) molecules was designed [8]. While novel in their own particular ways, these library based methods are labor and material intensive and lack the design element that will eventually lead to more reproducible and rational mimetic construction.

Mimetics of antibody combining sites represent a particularly interesting target. All antibodies have six CDRs residues all of which are more or less involved in antigen recognition. Thus, the antibody combining site is a predictable, if demanding example of the 'mimetic of a discontinuous surface.' Approaches to date have tended to simplify the problem by targeting the mimetic design to CDR H3, since this CDR is typically at the center of most, if not all, antigen interactions [2]. This has its limitations, however, since clearly other CDRs play an important role in the recognition process. The problem to now has been, how to incorporate residues from different CDRs into a single, synthetically accessible molecular design.

In this paper we described a new approach where a mimetic of the paratope of an antibody was designed in such a way that all the amino acids defined as 'active chemical groups' for the binding activity, and deriving from multiple CDRs, were incorporated into a specific, synthetically accessible, peptide-like construct. The monoclonal antibody (MAb) ST40 was selected to develop this method. This antibody is specific for the CDR3-like loop in domain 1 of the CD4 molecule and, indirectly, inhibits human immunodeficiency virus type 1 (HIV-1) replication.

The starting point for this process is a 3D structure of the target antibody and if possible complexed with its antigen. If no X-ray structure is available (as with ST40), a structure prediction method is required for the antibody and alternative methods like site directed mutation or alanine scanning [9] are required to identify the 'active chemical groups.' In this study, the 3D structure of ST40 was modeled using the research version of the AbM software [10] and the 'active chemical groups' were identified by alanine scanning [11] of synthetic overlapping peptides derived from the ST40 sequence.

Materials and methods

ST40 and alanine scanning

The cloning of MAb ST40 and its alanine scanning is described in [11].

CD4

The human recombinant CD4 used for the kinetic analysis and the biological assays was obtained from RepliGen (USA).

BIAcore analysis

The kinetic parameters, association rate constant (k_a) and dissociation rate constant (k_d), were determined by surface plasmon resonance (SPR) analysis using BIACORE 2000 (Biacore AB, Uppsala, Sweden). k_a and k_d were determined using BIAevaluation 3.0 software (BIAcore AB). The apparent equilibrium constant K_D is the ratio k_d/k_a. All experiments were carried out at 25 °C. The free NH_2 group of the lysine side chain of the mimetic was used to chemically immobilize molecules on a B1 sensor chip (BIAcore AB) following a standard EDC/NHS procedure from BIAcore. The SPR signals for immobilized peptide mimetics were found to be about 280–500 resonance units (RU) after completion of the chip regeneration cycle, which corresponds to 280–500 pg/mm^2. The binding kinetic of CD4 to immobilize the mimetic was determined by injecting several concentrations of CD4 (50–200 nM) in HBS buffer (running buffer) at a flow rate of 30 μl/min. For the selectivity study of the mimetic, the binding kinetics of immobilized mimetic were determined by injecting irrelevant proteins: Troponin C and 2C2 an anti-digoxin mAb, each at 165 nM, in HBS buffer at a flow rate of 30 μl/min. For the competition study, ST40 MAb (660 nM) and CD4 (165 nM) were pre-mixed, then co-injected on the sensor chip.

Protein binding analysis

The peptides were incubated for 10 min at 37 °C, at various concentrations (7, 15, 31, 61.5, 125, and 250 μM) in cell culture medium containing 10% FCS (fetal calf serum). Following incubation, the samples (450 μl) were centrifuged for 8 min at 11,000 rpm through an ultrafiltration membrane (Amicon Microcon 10 kDa) to separate the bound and free fractions of peptide. Fifty microlitres aliquots of ultrafiltrate were then analyzed by RP-HPLC. Results are expressed as percent free peptide in culture medium.

Anti-viral assay

Cells and HIV virus. Ficoll hypaque-isolated peripheral blood mononuclear cells (PBMCs) were obtained from a healthy donor and PHA-activated for three days. Cells were cultured in RPMI 1640 medium (Roche Products, Mannheim, Germany) supplemented with 20 IU/ml recombinant interleukin-2 (Roche Products), 10% fetal calf serum (FCS, Roche Products), 2 mM L-glutamine (Roche Products), and a 1% penicillin, streptomycin, and neomycin mixture (LifeTechnologies, Grand Island, USA) to a density of 2×10^5 cells/well in a 5% CO_2 atmosphere. Viral stocks (HIV-1_{LAI}) were prepared from PHA-activated umbilical blood mononuclear cell supernatants in the Neurology Service of the Commissariat à l'Energie Atomique (Fontenay aux Roses, France) and kept frozen at −80 °C until use. Fifty percent tissue culture infective dose ($TCID_{50}$) was calculated according to the Kärber formula [12].

Mimetics and control molecules. The PM2, PM3 mimetics, and the CONT1 peptide were diluted in culture medium at concentrations ranging between 200 and 0.5 μM. The anti-CD4 mAb ST40 and the anti-Troponin I mAb 9E8, kindly donated by CNRS UMR 5094, were tested at concentrations ranging between 660 and 6.6 nM. The two other inhibitors, the anti-CD4 mAb Q4120 and Azidothymidine (AZT), were provided by SpiBio (Fontenay aux Roses, France). They were used at concentrations of 0.33–33 and 1–100 nM, respectively.

HIV-1 infection assay. PBMCs (2×10^5/well) were plated in 96-well microplates in culture medium and pre-treated for 1 h at +4 °C with the inhibitors. Pre-treatment with inhibitors was performed in 1% FCS culture medium in the experiment with low $TCID_{50}$ to avoid serum

interference. Cells were then exposed to 100 HIV-1_{LAI} $TCID_{50}$ except for the last experiment, where lower doses of $TCID_{50}$ were used. After incubation for 4 h at 37 °C, each well was washed twice with 100 µl of culture medium. Plates were transferred at 37 °C in a 5% CO_2 atmosphere to allow infection. After 7 days in culture, supernatants were recovered and frozen at −20 °C until use. All experiments were performed in triplicate or quadruplicate. The amount of virus produced by PBMCs was monitored by measuring the reverse transcriptase (RT) activity in the supernatants using RetroSysR RT detection kit (Innovagen, Lund, Sweden). During the experiments, cell viability was microscopically checked by using trypan blue exclusion dye. Neither decrease of the cellular concentration nor presence of cellular fragments or a modification of cellular morphology was observed.

Molecular modeling

Antibody modeling. Molecular modeling of the light and the heavy chain CDRs of the antibody was carried out with the research version of the antibody modeling software A*b*M [10] running on an O2 R5000 Silicon Graphics workstation. The L2, L3, and H1 loops were constructed using canonical Class 1 frameworks and a canonical Class 2 framework for H2, as defined in A*b*M. A new canonical class had to be defined for the unusually long L1 CDR loop of ST40 containing 15 amino acids. Four X-ray structures of antibodies (1ibg, 1mf2, 1acy, and 1ggc) with L1 of 15 residues were identified and superimposed. Each of these L1's had a similar conformation. On the basis of this, the following canonical class for L1 was defined and the X-ray structures missing in the database were added: [TI]-x(20)-C(1)-x(1)-A-x(3)-V-x(7)-S-x(1)-[MLI]-x(1)-W(1)-x(35)-F [10]. The H3 loop of 13 amino acids, which is too long to fit into any H3 classification [13,14], was built using the conformational search program CONGEN [15] implemented in A*b*M, combined with a 3D structural database search. The definition of the building blocks to generate H3 with CONGEN was modified several times to obtain four different models since the 'kinked' feature was not very well defined. The X-ray structure of the antibody 1ibg, containing an L1 and a H3 loop of 15 and 13 residues, was then used to model H3 of ST40 with the same conformation. This model of H3 shows a kink and an extended form. Modeling was achieved through simple mutations and the structure was fully minimized using the Tripos force field. Hydrogens were added to all models of ST40 using the Sybyl software (Tripos) and the models were minimized during 100 iterations with the conjugate gradient method to eliminate all small steric conflicts. The solvent accessible surface areas of ST40 amino acids were calculated on the 3D model of ST40 by the SALVOL program developed by R.S. Pearlman et al. and implemented in Sybyl.

Design of the mimetics and molecular dynamics simulations. All the molecules were visualized and modified using the Sybyl software running on an O2 R5000 Silicon Graphics workstation. For the modeling of the mimetics, CDRs were selected from the antibody, maintained in their original conformations, and then linked together with chemical linkers via appropriate side chains and minimized. Molecular dynamics (MD) simulations of the mimetics were carried out in water with the AMBER force field using the AMBER software (University of California) running on a O200 R10000 Silicon Graphics workstation using four processors. The input files for AMBER were prepared with the XLEAP module. The cross-linking bridge, –CO–NH–, between side chains was defined as an amide bond. The water box was calculated with the module SolvateBox with a box size of 8 Å. The water boxes contained between 2500 and 3500 water molecules for all the simulations and were generated with the WATBOX216 module. This corresponds to a Monte Carlo distribution of water with periodic conditions and constant pressure. The temperature was fixed at 300 K, the cut-off was set at 10 Å, the dielectric constant was set at 1, and the sampling frequency was 1 ps. Before running the dynamics, the solvent molecules were minimized and then the solute. The MD simulations were run for 100 or 300 ps. The first 20 ps of the dynamics were the heating period, during which the system was taken from 10 to 300 K, and it was then increased by 15 K every 1 ps.

Analysis of the MD simulations. The first 50 ps of the MD simulations were not taken into account during the analysis since this was taken as the equilibration time. First, the mimetics at different times during the dynamics simulations were superimposed onto the starting conformer, which was very close to the conformation in the antibody. This displays the deformation of the mimetics during the MD simulation and was carried out using the Sybyl software. The root mean square deviations (RMSDs) were calculated, using the starting conformer as a reference, for each ps of the MD simulation with the CARNAL module from the AMBER package.

Synthesis of mimetics

The syntheses of the mimetics and the control peptide were carried out in a stepwise fashion on an Automated Multiple Peptide Synthesis (AMS 422, ABIMED). HPLC analyses were carried out on a Beckman LC126 system, using Waters SymetryShield column RP18, 5 µm, 100 Å (150 × 4.6 mm) with buffers A: 0.1% TFA in water and B: 0.08% TFA in acetonitrile; gradient from 95% A to 100% B in 12.5 mn with a flow of 1.5 ml/mn. HPLC purifications were carried out on a Waters Prep LC 4000 system, using Waters PrepPak Cartridge C18, 6 µm, 60 Å (40 × 100 mm); gradient from A to 60% B in 60 mn with a flow of 20 ml/mn. Mass analyses were performed using a MALDI-TOF spectrometer (Voyager DE Elite, PE Applied Biosystems) with dihydroxybenzoic acid as matrix. The peptide synthesis was performed using a polyethylene glycol graft polystyrene support (Fmoc-PAL-PEG-PS resin, substitution: 0.41 mmol/g). The Fmoc amino acids were activated in situ by the activating reagents DIPCDI (diisopropylcarbodiimide) and HOBt (1-hydroxybenzotriazole) in DMF (*N*,*N*-dimethylformamide) with a standard fourfold excess. An acetylation step was carried out after each amino acid incorporation to cap possible remaining amine groups and to ensure the absence of deletion peptides. The Fmoc protecting groups were removed by a solution of piperidine in DMF (20%) prior to each coupling. After cleavage of the support and removal of side chain protecting groups by reagent B, a standard TFA-based cocktail (88% TFA, 5% phenol, 5% H_2O, and 4% TIPS [v/v]) and a post-cleavage work-up by ether precipitation were employed. The purity of crude linear peptide was checked by analytical HPLC and mass spectroscopy. If purity was over 85%, the peptide was cyclized at room temperature using 1 equivalent of PyBop and 5 equivalents of $NaHCO_3$ in DMF. The reaction mixture was purified by preparative HPLC and the pure product lyophilized to obtain a white powder with a global yield from 5% to 20%. If purity of the linear peptide was less than 85%, an intermediate purification by preparative HPLC followed by lyophilization was achieved before cyclization. In the case of the mutated mimetic PM2, carrying an isoleucine to lysine substitution, the initial ε-NH_2 function of the lysine was specifically protected by the Alloc (allyloxycarbonyl) protecting group that was removed after the cyclization step according to the following conditions: $Pd[PPh_3]_4$ (3 equivalents), ACN/DMF/AcOH/NMM (7/2/2/1), r.t., 3 H. The pure products were analyzed by analytical HPLC and MALDI-TOF mass spectrometry.

Results and discussion

Modeling of ST40

The amino acid sequence of ST40 and the definition of the CDRs in A*b*M are shown in Fig. 1. ST40 contains two unusually long CDR loops, L1 (15 residues) and H3 (13 residues). The new canonical class identified by us for L1 permitted the construction of this loop in a spe-

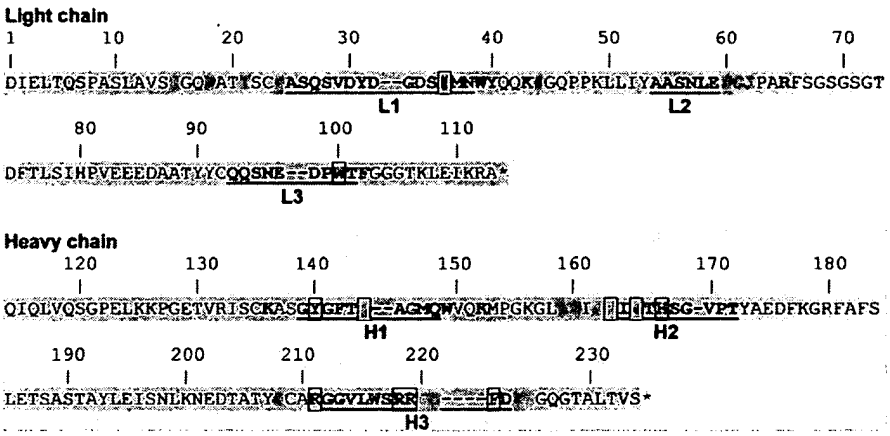


Fig. 1. Amino acid sequence of the variable regions of ST40 antibody. The CDR loops are underlined and bold. The important (red, bold) and less important (orange and yellow, bold) amino acid residues are represented. The residues selected to be included in the mimetic are boxed.

cific conformation, while for the H3 loop several possible conformations were identified. Only four X-ray structures of antibodies containing H3 loops of 13 residues have been described, each having a different conformation. A comparison of the 3D structures of H3 loops containing 11, 12, 13, 14, 15, and 16 residues was carried out and showed that the highest variability occurs at the apex of the loop. However, when the sequence contains an Arg at position N-1 of H3 and Asp-x-Trp at the C-terminus, as in ST40, a kinked form [13,14] is observed at the C-terminal end of the loop. This has been observed in the antibody 1ibg, which contains an L1 and an H3 loop of 15 and 13 residues, respectively. The selected 3D structure of H3 in ST40, containing an extended terminal part of the loop and a kink near the C-terminus, is shown in Fig. 2. This structure was selected as the reference model for the design of ST40 mimetics. However, the inherent flexibility of these longer H3 loops suggested that, during the mimetic construction, care should be taken not to constrain this region of H3 too heavily. Recent advances in the modeling of H3 regions were described by Whitelegg and Rees [16].

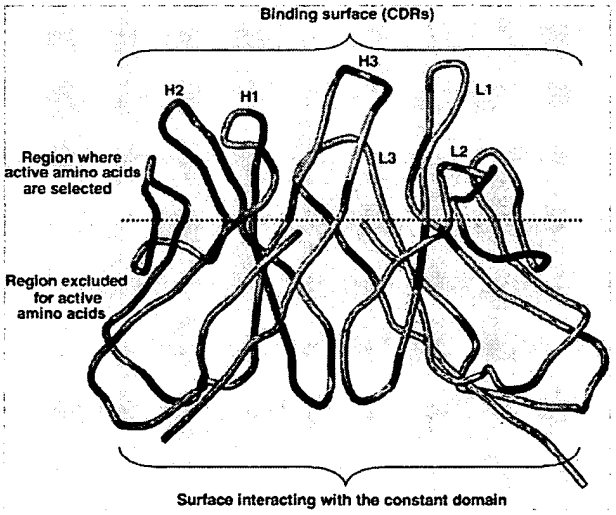


Fig. 2. Ribbon representation of the 3D structure of ST40 and localization of the important (red) and less important (yellow) amino acid residues, as determined by Alanine Scanning. The cut-off for residues with or without accessibility to the antigen is indicated by the dotted line.

Selection of the active chemical groups

The important amino acid residues of ST40, defined by peptide scanning followed by alanine scanning by Monnet et al. [11], are shown in Figs. 1 and 2. It is observed that some important amino acids are not located within, or in some cases even near, the CDR regions where the paratope is located (Fig. 2). To select the amino acids that should be included in the mimetic, the bioactive residues were displayed on the 3D structure of ST40 (Fig. 2) and their accessibility to water was calculated. Some residues were found to be located at some considerable distance from the paratope binding surface (top of Fig. 2), such as Lys152 and Lys43, whereas others were buried in the protein with poor accessibility to antigen. For example, many of the aromatic residues such as Tyr40, Phe102, or Trp149, often playing a structural role in all antibody variable regions, are amongst the buried residues identified as important for antigen binding by the mapping procedure. In contrast, charged residues such as Arg218 and Arg219, more likely to be directly involved in the binding, were conserved. In addition to these hydrophilic residues, amino acids that were predicted to be essential for maintaining the hydrophobic stabilization of the mimetic and which were likely to exhibit increased accessibility in the mimetic because of the flexibility of the H3 loop (e.g., Phe222, Trp163, and Tyr36) were conserved, even though their measured accessibilities on the static

model may have been low. Finally, a limit was fixed for those amino acid residues that should be included in the mimetic based on their importance as defined by the alanine scanning protocol, and their likely accessibility to the antigen. The amino acids selected are boxed in Fig. 1 and displayed on the loops in Fig. 3. All CDRs except L2 are represented in the mimetic by at least one residue. A framework residue located just before the H3 was also included in the mimetic (Arg211). Alanine scanning is a method which might generate false positive "active chemical groups" by exposing residues normally buried in the antibody paratope. As a result, even after careful selection, some of the amino acids selected to be part of the mimetic may not be necessary for binding. However, the first step of this study was to obtain an active mimetic and additional residues should not affect the binding site in the protein. A second step, not targeted in this study, would be the identification of each important residues of the peptide by mutation, to reduce the size of the mimetic. Crystallographic structures of antibody–antigen complexes would provide the optimal starting point for such studies.

Design of cyclic mimetics

To obtain a continuous peptide mimetic amenable to synthesis, the selected amino acids from the CDRs were extracted. The relative locations and orientations in space of the selected amino acid residues, especially the side chains, were conserved and the amino acids were linked together using glycine residues to obtain a cyclic peptide. The cyclization was performed by a covalent bond between the NH_2 of an internal lysine side chain and the C-terminal carboxyl of the peptide. This position was selected to ensure some flexibility within the N terminal segment of the mimetic, which contains the two arginines derived from the flexible H3 loop of ST40. To stabilize the mimetic the glycine residues were then replaced by hydrophobic residues at specific positions, or by proline residues for the turns, to obtain the most stable mimetic in which the important amino acids exhibited similar orientations as observed in the intact parent antibody, ST40. The selection of linking residues was critical for the design. Where possible, residues adjacent to 'active' residues were selected to mimic the environment in the parent antibody and to act in a packing role. For example the Tyr224 was used as a linker for its structural role.

After minimization and short MD simulations (100 ps) of the mimetic, the positions and orientations of the side chains of the important residues were checked. The short MD simulation allows any 'strong' deformation of the mimetic to be observed. At the end of the simulation the conformation of the mimetic was significantly perturbed compared to the starting conformation, and was more energetically favourable. The most stable mimetic from the MD simulations, having the lowest RMSD (4.4 Å) after 100 ps, was selected for synthesis. The amino acid sequence of this mimetic, named PM1, is shown in Table 1 and its starting 3D structure before MD simulation is shown in Fig. 3.

During the first 100 ps of MD simulation of the mimetic PM1, the majority of side chains moved less than 2 Å from their initial positions with the exception of Arg (long side chain) and the aromatic patch containing Phe222, Tyr36, and Trp163 (driven by intra-molecular interactions). After 100 ps, Arg218 and Arg219 within the flexible H3 loop remain oriented at the mimetic surface, pointing outwards into the solvent, as in the starting conformation. The distances between His167 Cβ and Arg219 Cβ, Tyr140 Cβ and Trp163 Cβ, and Asn144 Cβ and Trp100 Cβ are 17, 17, and 18 Å respectively. These are close to those seen in the starting conformation and suggest that global movements within the mimetic maintain the approximately correct distances between critical amino acids from the loops H2–H3, H1–H2, and L3–H1. When the simulation is extended to 300 ps the RMSD increases to 5.49 Å (average of the last 100 ps) due to the inherent flexibility of the peptide. While some deviation of residues from their positions in the original paratope is seen as the simulation progresses, a number of critical residues (His167, Arg218–219, Asn144, and Tyr36) remain in a favorable orientation. Due to problem of interaction of the PM1 mimetic with the sensor ship of the BIAcore, to determine its binding to CD4, the mimetic has to be fixed to the sensor chip using specific functionality. Therefore, a lysine residue was added to the sequence to facilitate the linkage of the mimetic on the sensor chip. Based on the average structure of the final 100 ps of the 300 ps dynamics simulation of PM1, the Ile between Tyr140 and Asn144 was identified as the best location for mutation. At this position the Lys side chain was observed to be pointing away from the surface believed to be important for the CD4 interaction. This change created mimetic PM2 (see Table 1). To demonstrate the importance of the structure of the mimetic for binding to CD4, a non-cyclized mimetic (PM3), having the same sequence as PM2, was synthesized (see Table 1). MD simulations of 300 ps duration were carried out for the peptides PM2 and PM3 using the same starting conformation as for PM1. Surprisingly, the non-cyclized peptide PM3 has a similar 'conformational profile' area as PM2 and maintains also a "compact" structure, probably due to the constraining influence of the five prolines. Indeed, the cyclization of PM1 and PM2 was straightforward during the synthesis. Therefore, as a final control, the peptide CONT1, in which all the linking amino acids of PM3 were mutated to Ala to de-structure the peptide, was synthesized (see Table 1). As expected, during the MD simulation, CONT1 exhibits a completely different conformational distribution—its starting cyclic conformation slowly disappeared during the simulation.

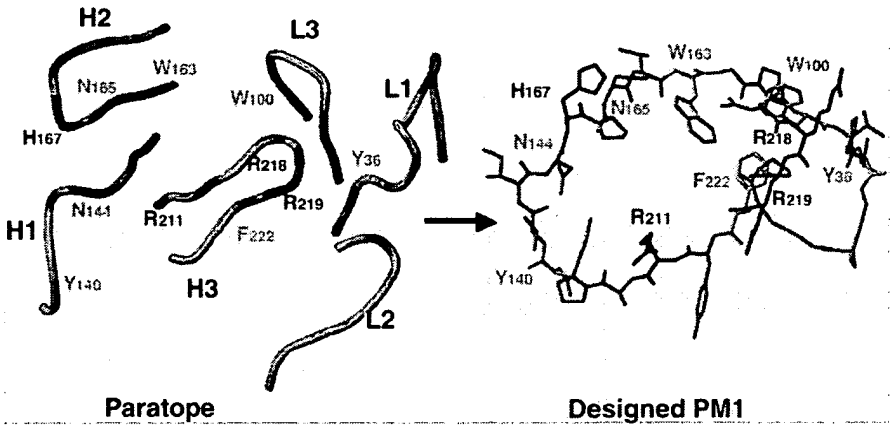


Fig. 3. Representation of the paratope of ST40 and of the resulting mimetic PM1 at 0 ps. The color coding for the important amino acids for inclusion in the mimetic is: red for charged residues (Arg and His), orange for Asn, and yellow for hydrophobic residues (Trp, Phe, and Tyr). The other amino acids of the CDRs are in cyan and the non-CDR residues of the antibody are not shown.

Table 1
Amino acid sequences of the synthesized mimetics

Name	Amino acids sequence
PM1	AcNHA$R_{218}R_{219}$PKF_{222}YR_{211}APY$_{140}$VI$N_{144}H_{167}$PN_{165}VW_{163}GPW_{100}VAY$_{36}$GP—NHCO—
PM2	AcNHA$R_{218}R_{219}$PKF_{222}YR_{211}APY$_{140}$VK$N_{144}H_{167}$PN_{165}VW_{163}GPW_{100}VAY$_{36}$GP—NHCO—
PM3	AcNHA$R_{218}R_{219}$PKF_{222}YR_{211}APY$_{140}$VK$N_{144}H_{167}$PN_{165}VW_{163}GPW_{100}VAY$_{36}$GPCOOH
CONT1	AcNHA$R_{218}R_{219}$AAF$_{222}$AR_{211}AAY$_{140}$AK$N_{144}H_{167}$AN_{165}AW_{163}AAW$_{100}$AAY$_{36}$AACOOH

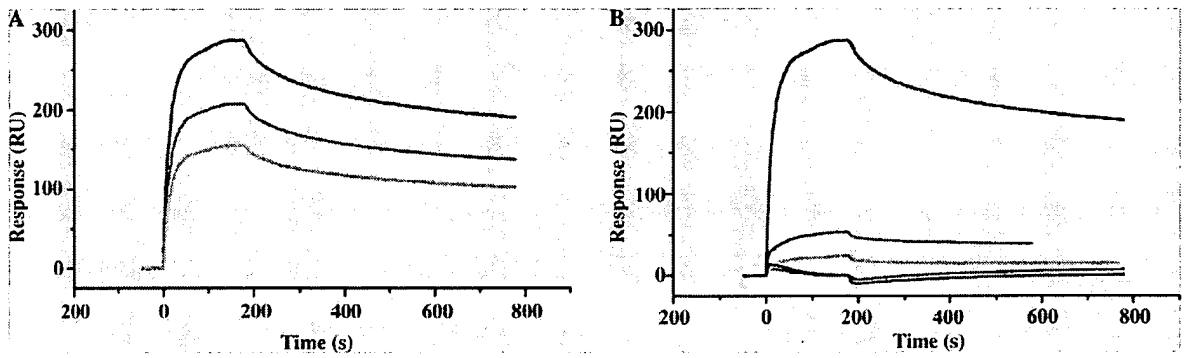


Fig. 4. (A) Sensorgrams of CD4 (50 nM, light blue; 99 nM, green; and 198 nM, red) binding to peptide PM2 immobilized on the sensor chip. (B) Sensorgrams of binding competition by injection of pre-incubated CD4 + ST40 (green) and CD4 only (red) when the mimetic PM2 is immobilized on the sensor chip. The binding of CD4 on CONT1 (blue) and to the irrelevant proteins 2C2 (pink) and TnC (dark blue) to PM2 is also represented.

BIAcore analysis

BIAcore sensorgrams for the binding of soluble CD4 to immobilized mimetic PM2 are shown in Fig. 4A. Kinetic constants were measured and gave values of $k_a = 10.4 \times 10^5\ M^{-1}\ s^{-1}$ and $k_d = 0.94 \times 10^{-4}\ s^{-1}$. The calculated K_D was 0.9 nM. The interaction of three irrelevant proteins (2C2, an anti-digoxin Mab, Troponin

C, and albumin) with the immobilized mimetic was also tested and the results are shown in Fig. 4B. No binding was observed for control proteins, except with albumin, which interacts with many peptides. For the competition study, ST40 mAb (660 nM, 50 μg/ml) and CD4 (165 nM, 20 μg/ml) were pre-mixed, and then co-injected on the sensor chip (Fig. 4B). The subsequent binding of CD4 to the mimetic was inhibited by 70%. This demonstrated that the mimetic PM2 binds to the same, or very closely located, site on CD4 as the parent antibody, ST40. A competition assay, with the CD4 fixed on the sensor ship and the peptide and ST40 co-injected, would have been more appropriate to demonstrate the binding competition, but the stickiness of the mimetic to the sensor ship did not permit us to perform this experiment.

The binding of PM3 to CD4 was not perturbed by the absence of covalent cyclization and its kinetic parameters were similar to those of PM2 (graphical data not shown) ($k_a = 9.03 \times 10^4\ M^{-1}\ s^{-1}$ and $k_d = 1.69 \times 10^{-4}\ s^{-1}$) with a calculated K_D of 1.87 nM. As for PM2, no irrelevant protein binding was observed, while pre-incubation of CD4 with ST40 inhibited 80% of CD4 binding to PM3 (Fig. 4B). As expected, the peptide CONT1 showed no binding to CD4 (Fig. 4B). These experimental results are entirely consistent with the MD simulation analysis of the PM2, PM3, and CONT1.

The affinities of PM2 and PM3, 0.9 and 1.9 nM, respectively, compared to 0.37 nM for the antibody [11], are very promising results for an antibody mimetic and lift these molecules into the range required for in vivo therapeutic activity.

Anti-retroviral activity assays

The ability of the anti-CD4 mimetics to inhibit viral reverse transcriptase activity was measured in PBMCs infected with 100 HIV-1_{LAI} $TCID_{50}$.

Protein binding of the peptides PM2, PM3, and CONT1 was measured using a 10% SVF culture medium (identical to that used in the cellular assays). After ultrafiltration using Microcon filters, 16 μM of PM2, for example, was found in the ultrafiltrate for an initial concentration of 125 μM. When all the results were averaged, the measured protein binding for the three compounds was: 85% of PM2, 82% of PM3, and 89% of CONT1. The observed non-specific 'stickiness' in a solution of PM2 and PM3, possibly originates from the hydrophobic 'underbelly' of the construct.

In the viral assay, the non-specific antibody 9E8 mAb showed no inhibitory activity, in contrast to AZT and Q4120 mAb, which inhibited RT activity in a dose-dependent manner (Table 2). Culturing PBMCs with the PM2 mimetic prior to infection with HIV led to the inhibition of viral reverse transcription, as also demonstrated for the parental anti-CD4 mAb ST40. The inhibitory activity was dose-dependent, with an IC_{50} for PM2 of ~7 μM (after correction for non-protein bound mimetic). The IC_{50} of the parental mAb was ~20 nM. Neither the negative control antibody 9E8 nor the CONT1 mimetic showed any dose-dependent activity, as expected. Higher concentration of PM2 was not tested due to problem of precipitation of the peptide in the presence of serum.

Table 2
Inhibition of viral particle production by anti-CD4 mimetic PM2

Inhibitor	Concentration	Inhibition of RT activity (%) first series	Inhibition of RT activity (%) second series
AZT	100 nM	100 ± 0	100 ± 0
	10 nM	74 ± 13	72 ± 17
	1 nM	66 ± 5	17 ± 17
9E8 mAb	660 nM	−5 ± 8	
	66 nM	13 ± 29	
	6.6 nM	−12 ± 11	
Q4120 mAb	33 nM	96 ± 0	
	3.3 nM	56 ± 16	
	0.33 nM	8 ± 11	
ST40 mAb	660 nM	96 ± 0	
	66 nM	83 ± 3	
	6.6 nM	14 ± 11	
PM2 mimetic	100 μM		98 ± 3
	50 μM	51 ± 13	74 ± 34
	5 μM	23 ± 5	—
	0.5 μM	−3 ± 6	—
CONT1	50 μM		11 ± 35
	10 μM		22 ± 25
	5 μM		−7 ± 30

Viral production was followed by measuring RT activity in cell-free supernatant. PBMCs were pre-treated with various inhibitors, washed, and then exposed to 100 $TCID_{50}$ of HIV_{LAI}. The data have been calculated from three triplicate experiments (means ± SD).

Finally, in order to assess the ability of the PM2 mimetic to reduce the viral titer, as the control Q4120 antibody does, we measured RT activity after infection of PBMCs with low HIV-1_{LAI} doses, ranging between 16 and 0.8 $TCID_{50}$. No viral replication was observed in the four culture wells tested, corresponding to PBMCs pre-treated with 100 μM (15 μM of free mimetic) PM2 and infected with 8 $TCID_{50}$, whereas 11 out of 12 culture wells infected with 8 $TCID_{50}$, but not pre-treated with PM2, showed viral replication. No cell death was seen during any of the assays (data not shown).

The low HIV cellular inhibition activity of PM2 was surprising considering high binding to CD4. Should cross-linking, or other ancillary effects on CD4 molecules by ST40, and other neutralizing antibodies, be advantageous for preventing HIV infection, this might explain the lower activity despite an essentially identical affinity for CD4. Again, the cellular assay took place over a period of 7 days so that the high stability of an intact antibody in the cellular milieu compared with a

smaller peptide would be likely to affect the overall activity performance.

Conclusions

We have described here a new approach, which is still under development and by which many different antibodies, whose function may simply be blocking of a pharmacologically relevant protein surface interaction, might be replaced by a smaller molecule. When applied to therapeutically important antibodies, the smaller size, easier production, and lower cost of the synthetic mimetics may offer significant advantages compared to recombinant antibodies or antibody fragments. This process seems promising given the binding interaction obtained, but needs to be reproduced in several antibody/antigen interaction systems, including those where binding alone is the functional readout. Such an approach obviously would not be appropriate where the full antibody structure plays a role in stabilising the antigen.

Acknowledgments

We acknowledge Mr. P. Turi and Dr. C. Quétard (BIACORE, France) for their useful advice and assistance for some of the BIAcore assays. We thank Steve Searle and Nicholas Whitelegg for help with aspects of the antibody modeling. We also acknowledge the French Research Minister for financial support.

References

[1] C.O. Ogbu, M.N. Qabar, P.D. Boatman, J. Urban, J.P. Meara, M.D. Ferguson, J. Tulinsky, C. Lum, S. Babu, M.A. Blaskovich, H. Nakanishi, F. Ruan, B. Cao, R. Minarik, T. Little, S. Nelson, M. Nguyen, A. Gall, M. Kahn, Highly efficient and versatile synthesis of libraries of constrained beta-strand mimetics, Bioorg. Med. Chem. Lett. 8 (1998) 2321–2326.

[2] B.-W. Park, H.-T. Zhang, C. Wu, A. Berezov, X. Zhang, R. Dua, Q. Wang, G. Kao, D.M. O'Rourke, M.I. Greene, R. Murali, Rationally designed anti-HER2/neu peptide mimetic disables $P185^{HER2/neu}$ tyrosine kinase in vitro and in vivo, Nature Biotech. 18 (2000) 194–198.

[3] G. Tuchscherer, B. Dörner, B. Sila, B. Kamber, M. Mutter, The TASP concept: mimetics of peptide ligands, protein surfaces and folding units, Tetrahedron 49 (1993) 3559–3575.

[4] M. Mutter, S. Vuilleumier, A chemical approach to protein design-template-assembled synthetic proteins (TASP), Angew. Chem. Int. Ed. Engl. 28 (1989) 535–554.

[5] C. Vita, J. Vizzavona, E. Drakopoulou, S. Zinn-justin, B. Gilquin, A. Ménez, Novel miniproteins engineered by the transfer of active sites to small natural scaffolds, Biopolymers 47 (1998) 93–100.

[6] R. Li, V. Dowd, D.J. Stewart, S.J. Burton, C.R. Lowe, Design, synthesis, and application of a protein A mimetic, Nature Biotech. 16 (1998) 190–195.

[7] U. Reineke, R. Sabat, R. Misselwitz, H. Welfle, H.D. Volk, J. Schneider-Mergener, A synthetic mimic of a discontinous binding site on interleukin-10, Nature Biotech. 17 (1999) 271–275.

[8] F. Falcioni, K. Ito, D. Vidovic, C. Belunis, R. Cambell, S.J. Berthel, D.R. Bolin, P.B. Gillespie, N. Huby, G.L. Olson, R. Sarabu, J.-M. Guenot, V. Madison, J. Hammer, F. Sinigaglia, M. Steinmetz, Z.A. Nagy, Peptidomimetic compounds that inhibit antigen presentation by auto immune disease-associated class II major histocompatibility molecules, Nature Biotech. 17 (1999) 562–567.

[9] D. Laune, F. Molina, G. Ferrieres, J.-C. Mani, P. Cohen, D. Simon, T. Bernardi, M. Piechaczyk, B. Pau, C. Granier, Systematic exploration of the antigen binding activity of synthetic peptides isolated from the variable regions of immunoglobulins, J. Biol. Chem. 272 (1999) 30937–30944.

[10] S. Searle, J. Pedersen, A. Henry, D.M. Webster, A.R. Rees, Antibody structure and function, in: C.A.K. Borrebaeck (Ed.), Antibody Engineering, Oxford University Press, New York, 1995, pp. 13–51.

[11] C. Monnet, D. Laune, J. Laroche-Traineau, M. Biard-Piechaczyk, L. Briant, C. Bes, M. Pugniere, J.-C. Mani, B. Pau, M. Cerutti, G. Devauchelle, C. Devaux, C. Granier, T. Chardes, Synthetic peptides derived from variable regions of an anti-CD4 monoclonal antibody bind to CD4 and inhibit HIV-1 promoter activation in virus-infected cells, J. Biol. Chem. 274 (1999) 3789–3796.

[12] G. Kärber, Arch. Exp. Path. Pharmak. 162 (1931) 956–959.

[13] H. Shirai, A. Kidera, H. Nakamura, Structural classification of CDR-H3 in antibodies, FEBS Lett. 399 (1996) 1–8.

[14] H. Shirai, N. Nakajima, H. Higo, A. Kidera, H. Nakamura, Conformational sampling of CDR-H3 in antibodies by multicanonical molecular dynamics simulation, J. Mol. Biol. 278 (1998) 481–496.

[15] R.E. Bruccoleri, M. Karplus, Prediction of the folding of short polypeptide segments by uniform conformational sampling, Biopolymers 26 (1987) 137–168.

[16] N.R. Whitelegg, A.R. Rees, WAM: an improved algorithm for modelling antibodies on the WEB, Protein Eng. 13 (2000) 819–824.

J. Mol. Biol. (1996) 262, 732–745

JMB

Antibody-antigen Interactions: Contact Analysis and Binding Site Topography

Robert M. MacCallum[1,2], Andrew C. R. Martin[1] and Janet M. Thornton[1,3]*

[1]*Biomolecular Structure and Modelling Group Department of Biochemistry and Molecular Biology University College London London WC1E 6BT, UK*

[2]*Graduate Programme MRC Laboratory for Molecular Cell Biology University College London, London WC1E 6BT UK*

[3]*Department of Crystallography Birkbeck College, London WC1E 7HX UK*

**Corresponding author*

We have analysed antigen-contacting residues and combining site shape in the antibody Fv and Fab crystal structures now available from the Protein Data Bank. Antigen-contacting propensities are presented for each antibody residue, allowing a new definition for the complementarity determining regions (CDRs) to be proposed based on observed antigen contacts. Contacts are more common at CDR residues which are located centrally within the combining site; some less central CDR residues are only contacted by large antigens. Non-contacting residues within the CDRs coincide with residues identified by Chothia and co-workers as important in defining "canonical" conformations.

An objective means of classifying protein surfaces by gross topography has been developed and applied to the antibody combining site surfaces. The surfaces have been clustered into four topographic classes: concave and moderately concave (mostly hapten binders), ridged (mostly peptide binders) and planar (mostly protein binders). We have determined the topographic classes for ten pairs of complexed and uncomplexed antibody-antigen crystal structures; four change topographic class on complexation.

The results will be of use in antibody engineering, antigen docking and in clinical immunology. To demonstrate one application, we show how the data can be used to locate the antigen binding pocket on antibody structures.

© 1996 Academic Press Limited

Keywords: molecular recognition; complementarity determining region; surface shape; antigen type

Introduction

The antibody repertoire is capable of recognising an almost infinite number of previously unencountered molecules, from small organic compounds to large macromolecular complexes. The structural basis for this recognition is surprisingly invariant; random genomic splicing of the light and heavy chain genes creates variability in sequence and length in just six loops (complementarity determining regions or CDRs) supported on a structurally highly conserved β-sheet framework. Improvements in binding affinity are achieved through somatic hypermutation of the CDRs and clonal expansion on exposure to antigen. The CDR residues number only ≈70 of the total ≈230 residues of the Fv antibody fragment (the smallest fragment able to bind antigen). Furthermore, the CDRs generally adopt only a limited number of canonical backbone conformations (Chothia *et al.*, 1989), determined by their loop length and certain key, "structurally determining" residues. Thus antibodies provide a system for the study of molecular recognition by proteins where structural variability is conveniently restrained, but where many examples of different inter-molecular interactions are available.

Since the first complexed antibody-antigen crystal structures were solved, numerous studies (Rees *et al.*, 1994; Webster *et al.*, 1994; Wilson & Stanfield, 1993, 1994, for reviews) of molecular recognition have focussed upon them. There are now (July 1995) 26 different complexed Fab and Fv structures in the Brookhaven Protein Data Bank (Bernstein *et al.*, 1977) and another 19 in uncomplexed form. In a recent review, Wilson & Stanfield (1994) have presented the variable contributions of each CDR to the antigen-buried surface, whilst Padlan and co-workers have

Abbreviation used: CDR, complementarity determining region.

0022-2836/96/400732-14 $18.00/0

© 1996 Academic Press Limited

published a by-residue summary of antigen contacts (Padlan *et al.*, 1995) with an emphasis towards sequence variability and humanization.

A number of groups have explored how detailed atomic interactions give specificity and affinity to antibody-antigen binding (Arevalo *et al.*, 1994; Wong *et al.*, 1995, for example). However there is also some interest in the broad-scale assembly of the combining site, and Webster *et al.* (1994) have suggested a subjective classification of antibody combining sites by general surface topography and antigen type. The objective computational analysis of combining site shape has so far been limited to just five light chain surfaces (Gerstein, 1992).

In this study we perform a detailed residue-level analysis of antigen contacts and investigate the contacts made by different sized antigens and different canonical loops. In addition we have developed an automated method for classifying surface topographies and support the suggestion that the general topography of antibody combining sites is correlated with antigen type. Finally, the application of our findings to the prediction of antigen contacting residues in uncomplexed antibody structures is discussed.

Results

Contact analysis

Solvent accessibilities and antigen contacts were computed for all residues in the 26 antibody-antigen complex structures (Table 1). Figures 1 and 2 summarize this information for each of the six CDRs. The bar graphs show how many antibody structures have solvent accessible or antigen contacting residues at each sequence position. The line graphs compare contacts made by different antigen types (see Figure legends for details).

The two main CDR definitions (Table 3) are shown in Figures 1 and 2 with arrows pointing towards the centre of the combining site (Figure 3). All residue numbering follows the scheme of Chothia *et al.* (1986; Chothia & Lesk, 1987).

Only two antigen contacts are made by residues outside those shown in Figures 1 and 2. These were L67 in antibody HyHEL-10 (3hfm, anti-lysozyme; Padlan *et al.*, 1989) and H64 in antibody NC41 (1ncd, anti-neuraminidase; Tulip *et al.*, 1992). Tables of the numeric data are available on the internet†.

Antigen contacts: CDR definitions

Not one of the residues in the combining site is in contact with antigen in all the structures (black bars in Figures 1 and 2). On average, potential antigen contacting residues form contacts in only 7.4 out of a possible 26 structures (28%) although at least one residue from each CDR forms contacts in 18 or more structures except in CDR-L2, which, as reported elsewhere (Wilson & Stanfield, 1993, 1994), only rarely contributes to antigen binding. Residues from CDR-L3 dominate; L91, L94 and L96 each make contacts to 21 different antigens. CDR-H3 is also dominant, but its length variability makes it impossible to assign special significance to particular residues.

In agreement with Padlan's survey of 22 complexes (Padlan *et al.*, 1995), the data show that for CDR-H1 and CDR-H2, where the loop and variability definitions differ, the common contacting residues correspond better with Kabat's sequence variability definition (Figure 2(a) and (b), broken arrows). In view of our data, however, we propose a new CDR definition based on observed antigen contacts (Table 3, column 4; see Table footnote for definition). In two cases (CDR-L1 and CDR-H2) the Kabat CDR definitions contain a large number of mostly non-contacting residues (L24 to L29 and H59 to H65, respectively) which can be explained simply by the gross arrangement of the CDRs with respect to the centre of the combining site, where most antigen interactions occur (Figure 3). A number of residues outside the standard CDR definitions make antigen contacts (number in parenthesis): L36 (1), L46 (2), L49 (6), H30 (5), H47 (4) and H93 (1). The residues making fewer contacts (L36, L46, H47 and H93) generally have less solvent accessibility (grey bars in Figures 1 and 2) and are often buried by other antibody residues. Because these residues are located in the centre of the combining site, they tend to make antigen contacts when they are exposed to solvent (CDR residues L34, L89, H35 and H50 also behave in this way).

Within each CDR (CDR-L3 and CDR-H3 in particular), more contacts are made at the end of the loop closer to the centre of the combining site, whilst apical loop residues interact less often with antigen.

Contacts and antigen type

Do the different types and sizes of antigen make use of different CDRs or CDR residues? We have divided the antibody-antigen interfaces into three groups based on interface size (see Methods and Data), but these also relate to antigen type as follows: small = haptens; medium = peptides, carbohydrates, nucleic acids; large = proteins, cyclosporin A (a cyclic peptide), and we generally refer to antigen classes by this size definition. The line graphs in Figures 1 and 2 show the differing contributions made by the three sizes of antigen.

As expected, we observe a preference for large antigens (thin continuous lines in Figures 1 and 2) to contact residues at the extremities of the combining site; for example, most of CDR-L2 and residues H30 to H32 in CDR-H1. Larger antigens also interact more with apical loop residues (for example, residues L92 to L94 in CDR-L3). Conversely at a number of central, non-apical CDR

† http://www.biochem.ucl.ac.uk/~martin/abs/GeneralInfo.html#contactdata.

Table 1. Antibody crystal structures (complexed with antigens; total 26), July 1995

PDB code	Name	Antigen	Antigen type	Interface area (Å²)[a]	Antigen size class[b]	Resolution (Å)/ *R* factor (%)	Topography[c]
2mcp	McPC603	Phosphatidyl choline	Hapten	32.63	Small	3.1/18.5	Moderately concave
1fig	1F7	TSA[d] for chorismate mutase (catalytic)	Hapten	54.97	Small	3.0/22.0	Moderately concave
1dbb	DB3	Progesterone	Hapten	71.61	Small	2.7/21.0	Concave
1eap	17E8	Phosphonate TSA (catalytic)	Hapten	76.46	Small	2.5/18.6	Concave
4fab	4-4-20	Fluorescein	Hapten	81.89	Small	2.7/21.5	Concave
1baf	AN02	DNP-spin-label	Hapten	82.40	Small	2.9/19.5	Planar
2cgr	2cgr	Cyanophenyl diphenylmethyl guanidineacetic acid	Hapten	87.73	Small	2.2/21.4	Concave
1igj	26-10	Digoxin	Hapten	91.19	Small	2.5/17.6	Concave
1ind	CHA255	Indium(3+)-eotube	Hapten	96.81	Small	2.2/18.8	Moderately concave
1ibg	40-50	Ouabain	Hapten	101.47	Small	2.7/20.9	Concave
1mfb	SE155-4	Heptasaccharide	Carbohydrate	118.88	Medium	2.1/16.0	Concave
1acy	59.1	HIV-1 GP120 peptide	Peptide	128.04	Medium	3.0/21.0	Ridged
1fpt	C3	Poliovirus type 1 peptide	Peptide	128.18	Medium	3.0/23.0	Ridged
1ggi	50.1	16aa peptide HIV-1 GP120	Peptide	135.47	Medium	2.8/18.8	Moderately concave
1him	17/9	8&9aa peptides from flu. HA	Peptide	137.23	Medium	2.9/20.0	Ridged
1cbv	BV04-01	ssDNA (auto-antibody)	DNA	139.19	Medium	2.7/19.1	Ridged
2igf	B13I2	Myohemerythrin residues 69–87	Peptide	146.65	Medium	2.8/22.0	Ridged
1tet	TE33	Cholera toxin peptide 3	Peptide	150.19	Medium	2.3/14.8	Ridged
1ikf	1ikf	Cyclosporin A	cyc-peptide	168.00	Large	2.5/16.4	Ridged
1jhl	D11.15	Pheasant egg lysozyme	Protein	174.42	Large	2.4/21.4	Planar
1jel	JE142	Histidine containing protein	Protein	191.50	Large	2.8/19.3	Concave
1vfb	D1.3	Hen egg lysozyme	Protein	193.05	Large	1.8/18.5	Planar
1mlc	D44.1	Hen egg lysozyme	Protein	198.26	Large	2.1/18.4	Planar
3hfm	HyHEL-10	Hen egg lysozyme	Protein	223.56	Large	3.0/24.6	Moderately concave
2hfl	HyHEL-5	Hen egg lysozyme	Protein	248.47	Large	2.54/24.5	Moderately concave
1ncd	NC41	Neuraminidase (N9 whale)	Protein	266.22	Large	2.9/15.7	Planar

[a] This is the solvent accessible contact surface area buried on complexing with antigen.
[b] Based on a simple clustering of interface area.
[c] Combining site topography classification.
[d] Transition state analogue.

Table 2. Antibody crystal structures (Fab/Fv only; total 19), July 1995

PDB code	Name	Antigen	Antigen type	Resolution (Å)/ *R*-factor (%)	Topography[a]
1fai	R19.9	Anti-arsonate (Xtal 2)	Hapten	2.7/18.9	Concave
1lmk	L5MK16	Anti-phosphatidylinositol scFv	Hapten	2.6/20.0	Ridged
2gfb	CNJ206	Catalytic (esterase)	Hapten	3.0/21.3	Concave
6fab	36-71	Anti-phenylarsonate	Hapten	1.9/20.9	Moderately concave
1gig	HC19	Influenza virus hemagglutinin	Peptide	2.3/19.5	Concave
1bbj	B72.3	Tumour marker (sialyl-Tn) (humanised)	Carbohydrate	3.1/?	Concave
2fbj	J539	Anti-galactan	Carbohydrate	1.95/19.4	Planar
1mam	YST9.1	Lipopolysaccharide A antigen of *Brucella abortus*	Carbohydrate	2.5/21.5	Concave
1bbd	8F5	Anti-HRV2	Protein	2.8/19.0	Ridged
1dfb	3D6	Anti-HIV-GP41	Protein	2.7/17.7	Concave
1fgv	H52	Humanised anti-CD-18 (some of H3 missing)	Protein	1.9/18.0	Planar
1for	Fab17-IA	Anti-human rhinovirus	Protein	2.75/17.4	Concave
1fvc	Hu-4D5	ERBB2 receptor extracellular portion (hum.)	Protein	2.2/18.3	Moderately concave
1rmf	R6.5	Anti-ICAM-1	Protein	2.8/18.8	Concave
1igc	MOPC21	Bound to *Streptococcus* protein G domain III	Unknown	2.6/16.8	Ridged
1igm	HuIgM	Human IgM Fv	Unknown	2.3/20.1	Planar
2fb4	Kol	Myeloma Fab	Unknown	1.9/18.9	Planar
7fab	New	Unsure	Unknown	2.0/16.9	Moderately concave
8fab	Hil	Human myeloma	Unknown	1.8/17.3	Moderately concave

[a] Combining site topography classification.

positions (L91, L96, H50) contacts are common with all antigen types. However, of the nine antigens contacted by residue H35, six are haptens (see previous section for a discussion of the solvent accessibility of this residue).

CDR backbone conformation: contacts in three-dimensional space

The non-contacting residues within the CDRs coincide with residues identified by Chothia *et al.* (1989) as important in defining canonical backbone conformations. Since, by definition, the side-chains of canonical key residues pack internally, they make few antigen contacts.

The variation in backbone conformations for the different canonicals is summarised in Figure 4. This highlights the broad conservation of different forms of CDR-L2, CDR-L3 and CDR-H2 and higher variability of CDR-L1, CDR-H1 and CDR-H3. In order to analyse the contacts made by the different CDR conformations, the mean fraction burial by antigen ($\bar{p}$, see below) is calculated for each CDR representative and is plotted in three dimensions in a "composite combining site" (Figure 4). This shows that, in general, only the central part of the combining site makes antigen contacts. The contact epicentre is displaced towards the heavy chain CDRs.

Antigen contacts made by specific backbone conformations can be compared in CDR-L1 and CDR-H2, where there are sufficient numbers of different examples (six or more). These two CDRs are shown in detail in Figure 5. Where CDR residues in two different canonical conformations are in the same position and orientation, the average contacts made are qualitatively similar. The insertion in CDR-L1 alters the side-chain orientations of residues L30 and L31 (Figure 5(a) and (b)) such that in the longer class 4 loop these never make contacts (in the shorter class 2 loop they do make contacts).

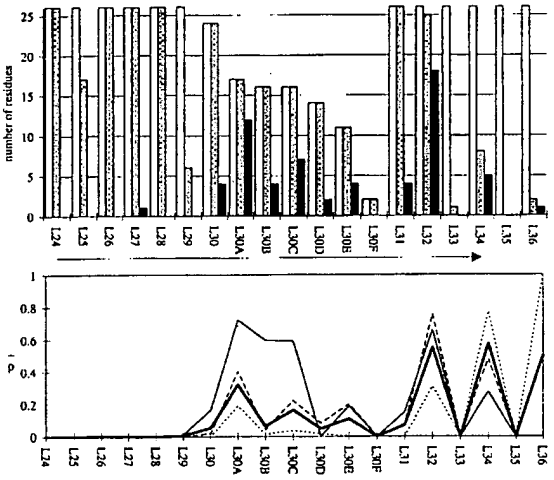
Antibody surface topography

To analyse patterns of surface topography, a fractal measure was used to identify the gross surface shapes which have, until now, been classified by eye: i.e. cavity, groove and planar (Webster *et al.*, 1994). The method reduces a three-dimensional molecular surface patch (of any size) to a two-dimensional composition profile which quantifies the relative amounts of concavity and convexity, and uses the fractal atomic density measure of Kuhn *et al.* (1992) to quantify local surface curvature. In Figure 6(a), the surface convexity/concavity composition profiles are shown for antibody-antigen interface surfaces; Figure 6(b) to (d) shows the same profiles for the whole combining site surfaces (using our "contact" CDR definition, see Table 3).

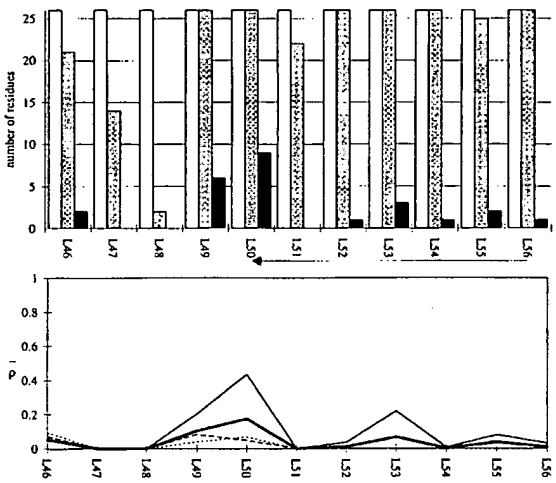
Antibody-antigen interface

It is generally accepted that the convexity/concavity of the combining site depends largely on the radius of curvature of the antigen, since the amount of antigen surface buried in complexes is always high and the surfaces are complementary (Wilson & Stanfield, 1993; Tulip *et al.*, 1992; Lawrence & Colman, 1993). When applied to the actual antibody-antigen interfaces (obtained using

1. (a) CDR-L1



(b) CDR-L2



(c) CDR-L3

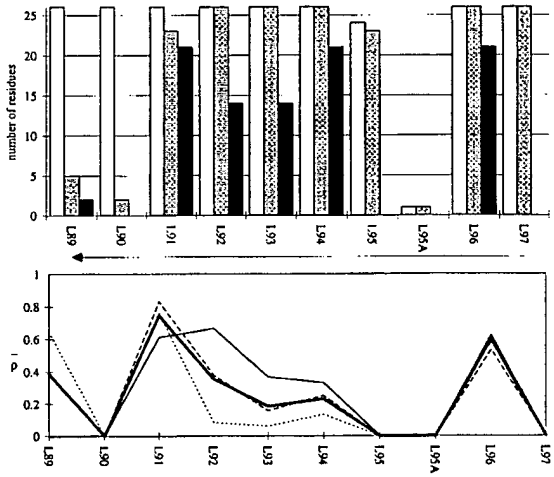


Figure 1. Contact analysis for light chain antigen contacts. Unfilled bars, number of antibodies with residues at each sequence position. Grey bars, number of antibodies with accessible residues at sequence position. Black bars, number of antibodies making antigen contacts at each position (see Methods and Data). Line graphs: $\bar{\rho}$, mean fractional burial by antigen type (see Methods and Data). Thick continuous lines, all antigens. Fine broken lines, small antigens (see Table 1, column 6). Medium broken lines, medium sized antigens. Thin continuous lines, large antigens. Arrows mark the standard CDR definitions (see Table 3). Where these differ, the broken arrows denote the Kabat sequence variability definition; the filled arrows denote the Chothia loop definition. Arrows point towards the centre of the combining site (see Figure 3).

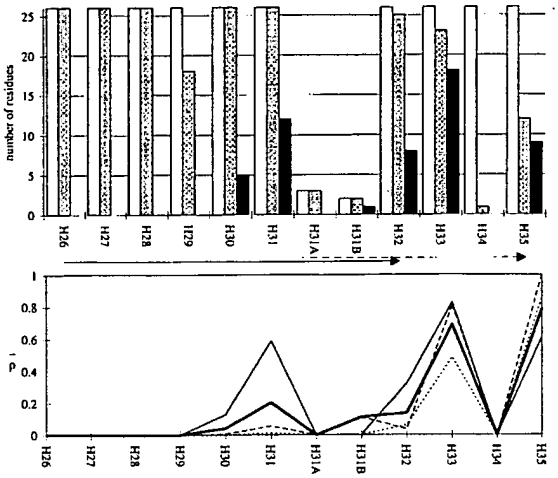
GRASP (Nicholls *et al.*, 1991) and defined as any antibody surface less than 2.8 Å from the antigen surface) our surface comparison (Figure 6(a)) clearly supports this; small antigens have the most concave interfaces, followed by medium and large sized antigens, respectively.

Whole combining site

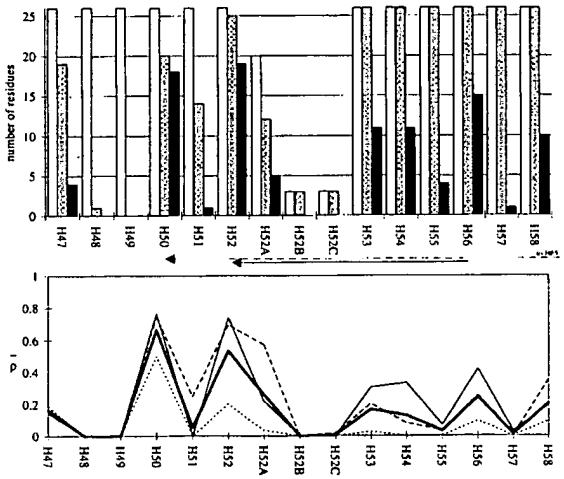
However, the trend is much less clear when the whole antibody combining site is subjected to the same analysis; Figure 6(b) presents the surface convexity/concavity composition profiles for the molecular surfaces overlying the "contact defined" CDR residues (Table 3) for the 26 complexed antibody crystal structures. A number of the anti-hapten antibodies (green lines in Figure 6(b)) have the highest proportion of concave surface points, but there is considerable overlap between the profiles of medium and large antigen binding sites. The loss of distinction between the antigen classes is caused by the increase in noise resulting from the inclusion of random, non-interface surfaces in the calculations.

In order to eliminate the subjectivity of these observations we have performed cluster analysis on the full set of 45 whole combining site surface topography profiles from both complexed and uncomplexed structures. Using Ward's minimum variance method (Ward, 1963; Wishart, 1969), the topography composition profiles (as multi-dimensional vectors) cluster into four subclasses of roughly equal size. Figure 6(c) presents the profiles for the complexed and uncomplexed antibodies; the colours refer to the cluster membership of each surface. The near-centroid representatives of each

2. (a) CDR-H1



(b) CDR-H2



(c) CDR-H3

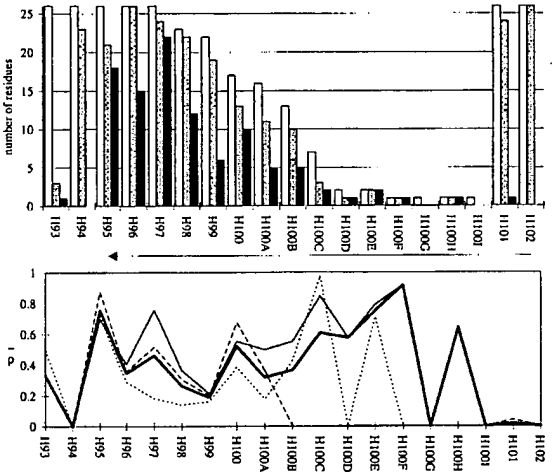


Figure 2. Contact analysis for heavy chain antigen contacts. See the legend to Figure 1. Note that in CDR-H2 the sequence variability definition extends to H65.

cluster are shown in Figure 6(d). Descriptions of each cluster are based on their general shape properties: planar, ridged, concave and moderately concave.

Ridged-type combining site surfaces are characterised by a large proportion of convex surface in addition to concave surface points. Note that our strictly compositional surface analysis cannot describe contiguous features, such as long and narrow surface depressions which one would describe as grooves. We therefore use the term "ridged" although for practical purposes this is equivalent to the established "groove" classification.

Tables 1 and 2 show which topographic class (i.e. cluster) is assigned for each antibody in our data set. Tables 4 and 5 show the distributions of topographies by antigen type. For the complexed data set (Table 4) there is a good, but not perfect, correlation between antigen type and surface topography, supporting the cavity-groove-planar classification (Webster *et al.*, 1994) in principle. While others have suggested three shape classes, our analysis shows that four classes fit the data better with both concave and moderately concave being predominantly hapten binders.

In an attempt to explain the occurrence of a number of outliers, including a planar anti-hapten antibody (AN02, Brünger *et al.*, 1991) and a concave-type anti-protein antibody (JE142, Prasad *et al.*, 1988), we assessed the relative contributions of interface and non-interface surfaces to the measure of combining site topography. Figure 6(a) highlights selected outliers from the whole combining site analysis (thick lines). The AN02 antibody-antigen interface has a fairly typical concave topography for a small antigen; its planar combining site classification arises because the

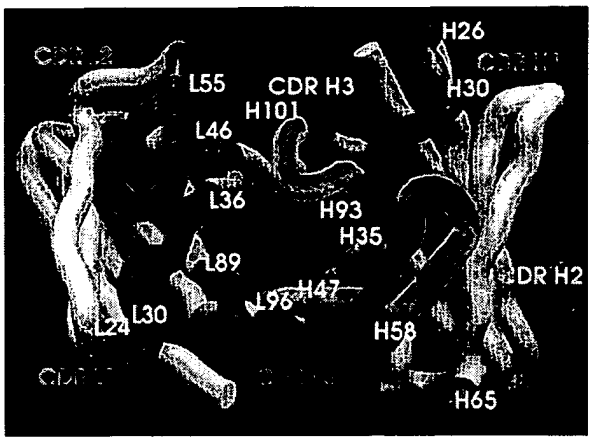


Figure 3. CDR arrangement. Contact CDR definitions (see Table 3) are shown in strong colours, residues defined by other CDR definitions are shown in grey.

Table 3. CDR definitions: all residue numbering follows the scheme of Chothia *et al.*

CDR	Kabat[a]	Chothia[b]	Contact range[c]
L1	L24–L34	L24–L34	L30–L36
L2	L50–L56	L50–L56	L46–L55
L3	L89–L97	L89–L97	L89–L96
H1	H31–H35	H26–H32	H30–H35
H2	H50–H65	H52–H56	H47–H58
H3	H95–H102	H95–H102	H93–H101

[a] Sequence variability (Wu & Kabat, 1970).
[b] Loop definition (Chothia *et al.*, 1986; Chothia & Lesk, 1987).
[c] Start and end residues must satisfy $\bar{\rho} > 0.01$ and make at least one antigen contact (see Methods and Data) in the set of complexes (see Table 1).

non-interface combining site surface is so flat. In contrast, the interface between antibody JE142 and its protein antigen has a relatively large amount of concave surface (compared with other large antigens) and its concave-type combining site classification (planar is expected) is therefore only partly influenced by non-interface surface topography.

Uncomplexed antibodies and "induced fit" movements

The topographies of the uncomplexed antibody combining sites (Tables 2 and 5) are less clearly correlated with antigen type. However, for many of these antibodies it has been more difficult to identify the true nature of the antigen and therefore to know which type of combining site to expect.

It is possible that a combining site "relaxes" in the absence of antigen (induced fit) or is distorted by crystal contacts. We have investigated the former possibility by determining the topographic classes of the combining site surfaces of both the complexed and uncomplexed antibody structures where available. The topographies of the ten pairs of complexed/uncomplexed antibody are given in Table 6. In these cases, the assigned topographic class is that of the closest topography profile in Figure 6(c) (minimum Euclidean distance). Four antibodies change their topographic class between complexed and uncomplexed forms; two of the ridged-type antibodies, 17/9 (Schulzegahmen *et al.*, 1993) and BV04-01 (Herron *et al.*, 1991), become concave-type, and two moderately concave-type antibodies (McPC603, Satow *et al.*, 1986; and 50.1, Rini *et al.*, 1993) become ridged-type. However, none of the concave-type anti-hapten antibodies relax to moderately concave, nor do the planar-type antibodies change their gross topography.

Predicting antigen contacts

Having investigated relationships between antibody sequences, antigen contacts, antigen type and combining site topography, we applied these results by predicting the antigen-binding pockets of antibody structures. This is particularly relevant for antibodies whose sequences are known and from which a model can be constructed. From the model we aim to predict those residues most likely to be involved in antigen contact, which can then be

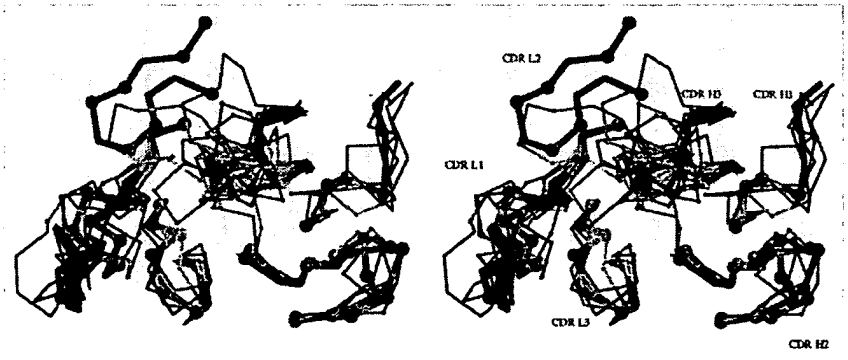


Figure 4. Antigen contacts and CDR conformation. Mean burial by antigen data ($\bar{\rho}$) are shown on representatives of each different CDR conformation found in the complexed antibodies. Colours range from blue (zero mean burial) to red (high mean burial). CDR conformations shown in thick lines have six or more examples in the data set. Space-filled carbon alpha atoms make at least one antigen contact. CDRs are fitted at framework atoms (see Methods and Data).

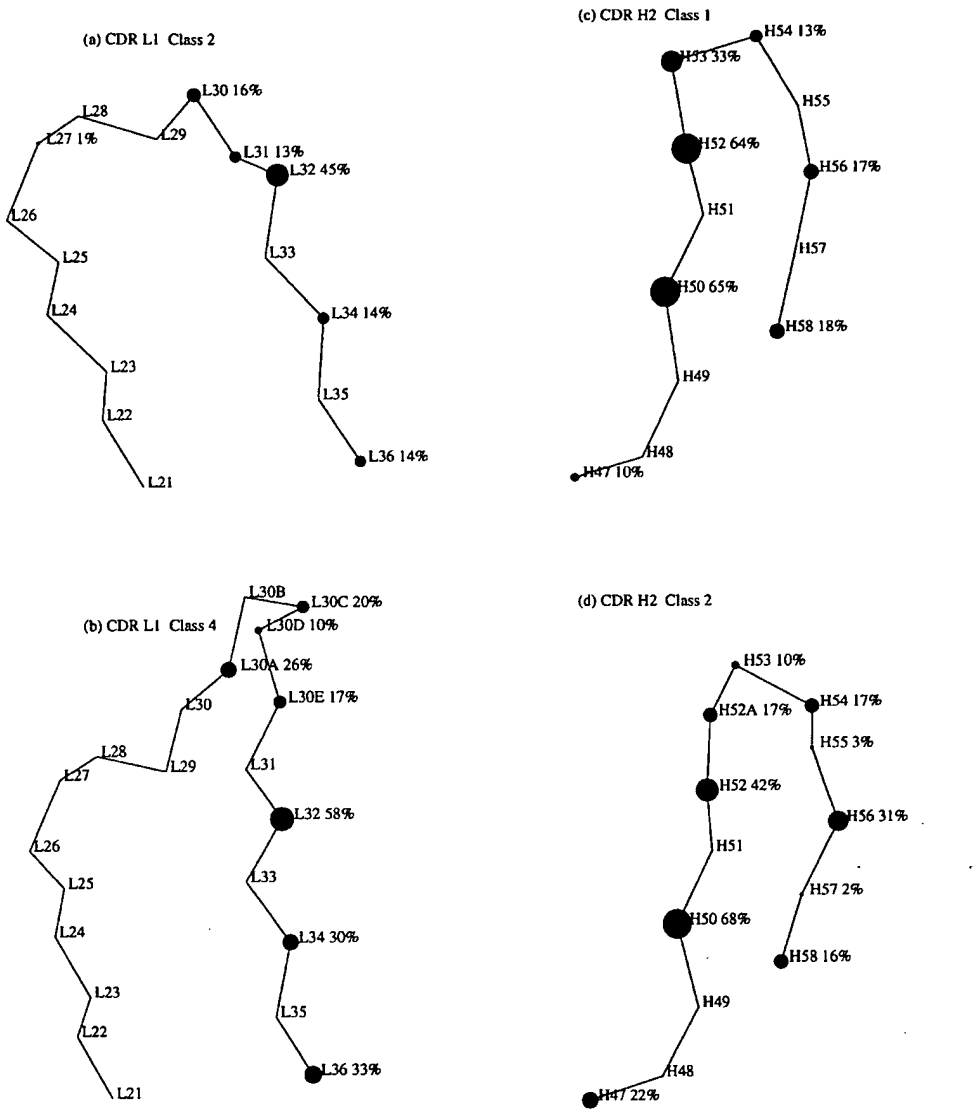


Figure 5. Comparing the antigen contacts made by different canonical loop conformations. (a) CDR-L1 canonical class 2 (7 examples). (b) CDR-L1 canonical class 4 (six examples). (c) CDR-H2 canonical class 1 (six examples). (d) CDR-H2 canonical class 2 (11 examples). Mean fractional burial by antigen ($\bar{\rho}$) is shown (as a percentage) for each antigen contacting residue. Carbon alpha surface area is also proportional to $\bar{\rho}$.

subjected to site-directed mutagenesis to test both theory and model.

We have developed a rapid and simple method which uses only average contact and surface shape information and test it by application to the complexed crystal structures (all antigen-contact data from the antibody being predicted are excluded). Firstly the mean burial (by antigen) data, $\bar{\rho}$, for each residue is projected onto the molecular surface of the antibody (Figure 7(a)). Figure 7(b) shows the same antibody with the surface coloured according to its curvature, C, as calculated by the program GRASP† (Nicholls *et al.*, 1991). This alternative description of curvature was adopted for purely practical reasons and allows easy application of the technique by anyone with access to GRASP. To find the probable contact surface we calculate $\bar{\rho}(30 - C)$, which combines the measure of concavity with the probability of antigen contact. An arbitrary cutoff of $\bar{\rho}(30 - C) > 35$ was chosen by visual inspection (using just one complex), and surface points satisfying this condition are coloured red (Figure 7(c)). This red surface is patchy and discontinuous, so neighbouring patches are merged and disconnected patches are eliminated using a dilation/erosion procedure (Delaney, 1992) developed within GRASP. The red surfaces are dilated five times, eroded eight times and dilated again

† GRASP macros are available from http://www.biochem.ucl.ac.uk/~bob/Ab/graspMacro.txt.

three times (each by 1 Å). The binding pocket prediction is the resultant smooth red patch (Figure 7(d)).

The "worst case" accuracy, Q_w, and "fraction correct" accuracy, Q_c, for predicting antigen contacting residues by this protocol with each of the 26 complexed crystal structures are calculated as follows:

$$Q_w = \frac{N_o + N_u}{N_c}$$

$$Q_c = \frac{N_c}{N_i}$$

where N_c is the number of correctly predicted antigen contacting residues, N_i is the total number of residues in the antibody-antigen interface and N_o and N_u are the numbers of over- and under-predicted contact residues, respectively. The mean Q_w is 1.0 (best 0.25, worst 2.5) whilst the mean Q_c is 0.79 (best 1.0, worst 0.53). As expected (since contact data are derived from all antigen classes), the method tends to over-predict interactions with smaller antigens, and under-predict interactions with larger antigens. Improvements can be made by using separate average contact data for different sizes of antigen. Human judgement can also play an important role in using these predictions. For example, Figure 7(e) shows the prediction for antibody 40-50 (1ibg, anti-ouabain, Jeffrey *et al.*, 1995); the choice between the two putative binding regions is simple if it is known that the antigen is small and could not bind across both predicted surface patches.

Discussion

Antibodies have been used in biomedical research for many years and have started to be used in applied medicine. They have potential roles in fields as diverse as cancer therapy (Chester & Hawkins, 1995), biosensors (Killard *et al.*, 1995) and

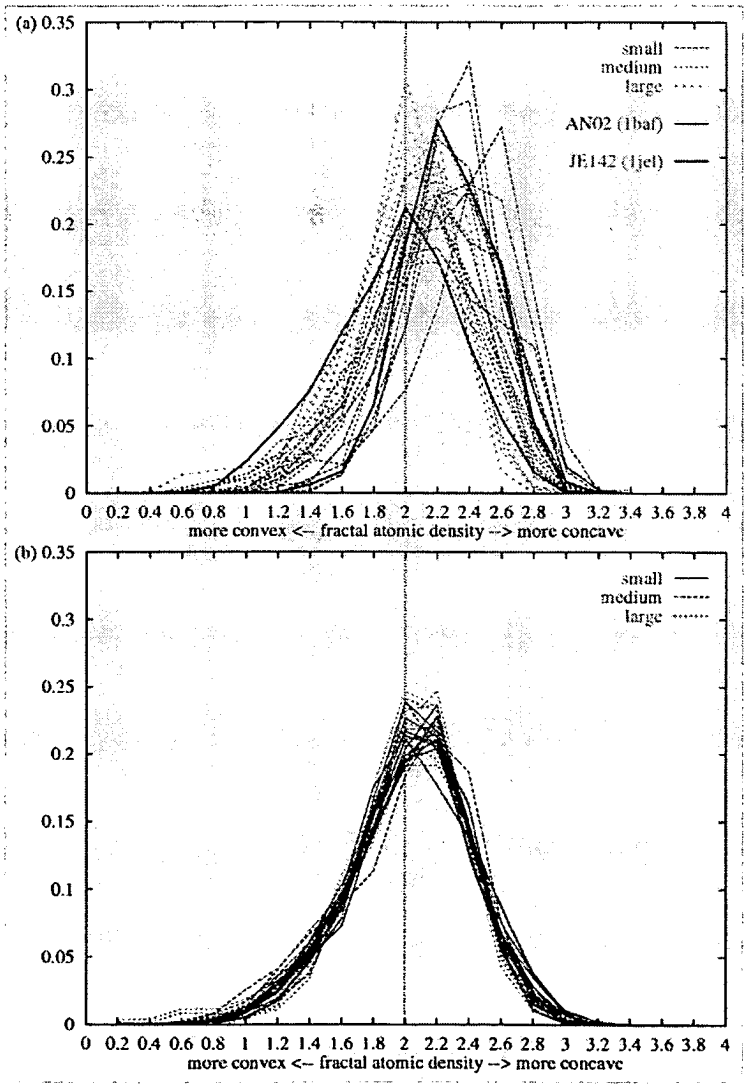


Figure 6(a) to (b) *legend overleaf*

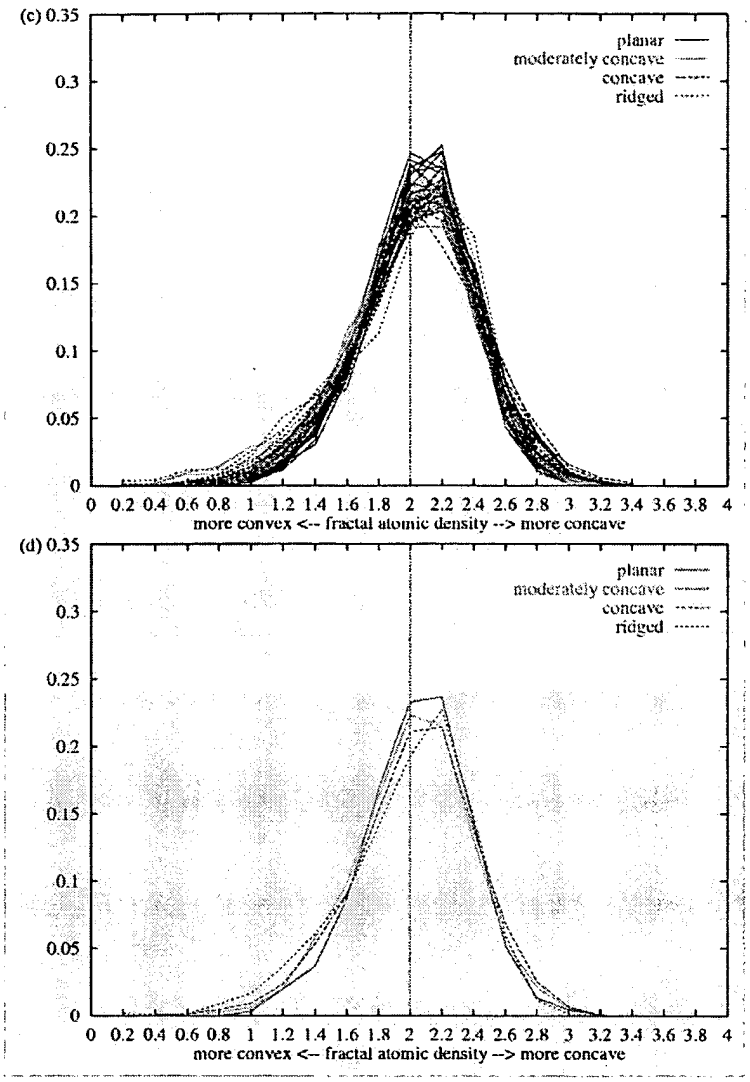


Figure 6(c) to (d)

Figure 6. Convexity/concavity composition profiles for antibody molecular surfaces. (a) Profiles for the 26 antibody-antigen interface surfaces. Data are grouped into antigen/interface size class (see Table 1 and Methods and Data). Small antigens have more concave binding pockets whilst larger antigens have flatter binding surfaces (more points with fractal atomic density ≈ 2.0). (b) Profiles for the whole combining site surfaces (defined as the surface overlying contact defined CDR residues, see the text) for the 26 complexes (colours are as in (a)). (c) After clustering, whole combining site surfaces (see (b)) for all 45 complexed and uncomplexed antibodies. Line colours and descriptions refer to the four clusters determined for these data (see the text). (d) Near-median representatives of each cluster in (c).

catalysis (Hilvert, 1994). They are also important molecules in autoimmune diseases such as rheumatoid arthritis and systemic lupus erythematosus (Brinkman *et al.*, 1990). Research and medical applications of antibodies have been made possible by recent advances in library-based *in vitro* production of monoclonal antibodies (Winter *et al.*, 1994), computer based structure prediction from primary sequence (Chothia *et al.*, 1989; Martin *et al.*, 1991), and the availability of an increasing number of experimentally determined three-dimensional antibody structures.

Analysis of the antibody combining site has four ultimate goals: (1) the ability to tailor the combining site to modify its binding or to introduce novel function; (2) *de novo* design of antibodies to target

Table 4. Distribution of antigen types within the four combining site shape classes for the complexed antibody structures (see also Table 1 column 8)

Antigen size	Planar	Ridged	Moderately concave	Concave
Small	1	0	3	6
Medium	1	6	1	1
Large	4	1	2	1

Table 5. Distribution of antigen types within the four combining site shape classes for the uncomplexed antibody structures (see also Table 2 column 6)

Antigen size	Planar	Ridged	Moderately concave	Concave
Small	0	1	1	2
Medium	0	0	0	3
Large	1	1	1	3
Unknown	2	1	2	0

[a] Small = hapten; medium = peptides/carbohydrates; large = proteins.

antigens; (3) identification of antigens to which antibodies bind; (4) the ability to dock antibody and antigen. Together with methods for modelling the antibody combining site, our analysis of combining site shape and contacting residues is applicable to all these areas.

The position and orientation of a residue relative to the centre of the combining site are the major determinants of its propensity to bind antigen. Our analysis of contact residues allows tailoring and design *via* mutagenesis experiments to concentrate on those residues which are most likely to interact with antigen. Correlations of CDR amino acid composition with antigen type are much stronger when the analysis is restricted to common contacting residues rather than the complete hypervariable surface (unpublished results).

As we have shown, contact residue analysis also allows us to identify the region of the antigen combining site most likely to interact with antigen and therefore restricts the area of the molecule which must be considered in docking experiments. In addition, mean burial data, $\bar{p}$, could be used as part of a potential function to score models in an antibody-antigen docking algorithm.

It has long been proposed that antibodies, binding to protein antigens, are characterised by a flat combining site, while those binding to peptide and DNA antigens have a groove-like combining site and hapten-binders a cavity (Webster *et al.*, 1994).

Our analysis of the shape of the combining site has shown that these broad shape classes of antibody combining sites can be identified using a fractal measure. However, the correlation with antigen type is not exact. In antibodies whose crystal structures are available, both complexed with antigen and uncomplexed, there is evidence for conformational change (Stanfield *et al.*, 1993). Our analysis has shown that whilst changes in the gross topography do occur, they are not of great magnitude (i.e. never from cavity to planar or *vice versa*). Therefore the shape analysis, used cautiously, should prove useful in the design of antibody combining sites for specific antigen targets and in the identification of potential antigen types for autoimmune antibodies.

In summary, we have shown that antibody-antigen interactions, although random and combinatorial in the origin of antibody diversity, are not completely chaotic. Antigens tend to bind to the antibody residues located at the centre of the combining site where the six CDRs meet and we have provided a contact definition of the CDRs which may prove useful in antibody design and mutagenesis. Our detailed analysis of combining site shape has largely confirmed the proposed correlation between shape and antigen type although we have identified a number of noteworthy outliers. Thus our analysis adds to the understanding of antibody-antigen interactions and the mechanisms by which antibodies recognise their targets.

Methods and Data

Antibody structures and sequences

The data set of antibody structures used in this study includes all the complete light/heavy chain dimers available in the Brookhaven Protein Data Bank (PDB; Bernstein *et al.*, 1977) January 1995 release and pre-releases up to July 1995, with three exceptions. Antibody 26/9 (PDB code 1frg, Churchill *et al.*, 1994) has been excluded since it is very similar (91% and 88% identity in the light and heavy chains, respectively) to 17/9 (1him, Schulzegahmen *et al.*, 1993) and has the same antigen. Also excluded are the pre-release structures for antibodies F9.13.7 (1fbi, Lescar *et al.*, 1995) and JEL103 (1mrc-f, Pokkuluri *et al.*, 1994) which have large numbers of missing atoms. Where more than one antibody structure was available, the highest resolution structure was chosen. All water molecules have been removed

Table 6. Combining site topographies compared between complexed and uncomplexed antibody crystal structures

Antibody name	Antigen type	Complexed PDB code	Complexed Topography	Uncomplexed PDB code	Uncomplexed Topography
DB3	Hapten	1dbb	Concave	1dba	Concave
26-10	Hapten	1igj	Concave	1igi	Concave
"2cgr"	Hapten	2cgr	Concave	1cgs	Concave
McPC603	Hapten	2mcp	Mod. concave	1mcp	Ridged
50.1	Peptide	1ggi	Mod. concave	1ggb	Ridged
17/9	Peptide	1him	Ridged	1hil	Concave
B1312	Peptide	2igf	Ridged	1igf	Ridged
BV04-01	DNA	1cbv	Ridged	1nbv	Concave
D1.3	Protein	1vfb	Planar	1vfa	Planar
D44.1	Protein	1mlc	Planar	1mlb	Planar

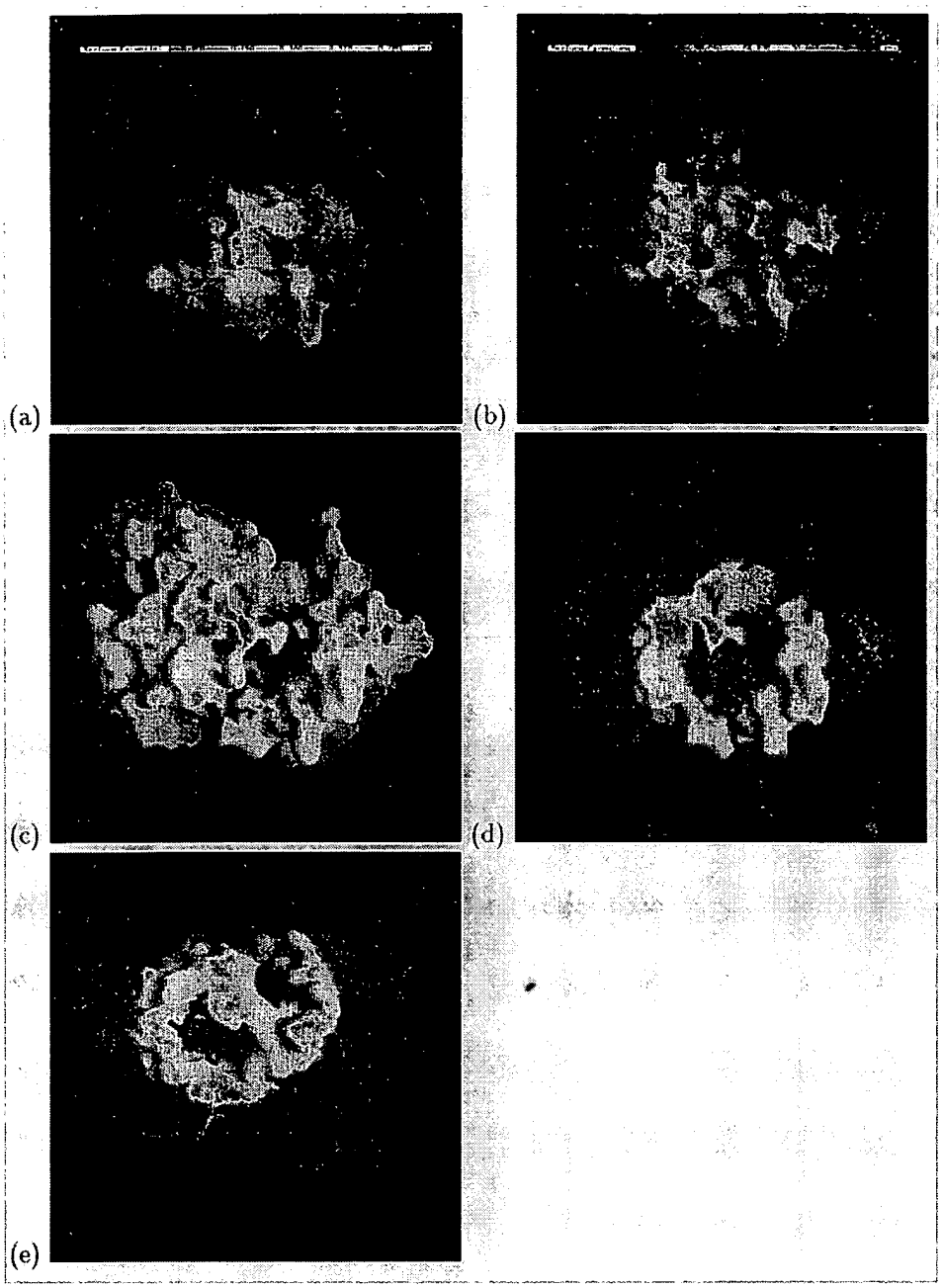


Figure 7. Antigen binding pocket prediction using mean fractional burial by antigen calculated from the set of 26 antibody-antigen complexes. Sequence (a) to (d) Antibody CHA255 (Love *et al.*, 1993) molecular surface: (a) coloured by predictive mean fractional burial by antigen, $\hat{p}$; (b) coloured by GRASP curvature, C; (c) red patches indicate where $\hat{p}(30 - C) > 35$; (d) final binding pocket prediction (red patches in (c) have been smoothed) and bound antigen (black wireframe). (e) Antibody 40-50 (Jeffrey *et al.*, 1995), prediction as in (d). Figures produced using the program GRASP (Nicholls *et al.*, 1991).

from the coordinate files. The data set is presented in Tables 1 and 2.

All sequences are numbered according to Chothia & Lesk (1987) using an alignment with a consensus sequence and the program *kabatnum* by ACRM†. This is not in accordance with the scheme used by Chothia *et al.* in their later papers (Chothia *et al.*, 1989, and more recent papers) where they place the insertion at L31. However, examination of the CDR-L1 conformations shows that L30 is the correct site for the insertion.

Three antigen size classes (Table 1 column 6) have been introduced in order to simplify the text. Although the classes were defined by simple cluster analysis of the antibody-antigen interface areas, they correlate with the antigen types as follows: small (haptens); medium

† http://www.biochem.ucl.ac.uk/~martin/PROGRAMS.TXT.

(peptides, carbohydrates, nucleic acids); large (proteins, cyclosporin A).

Accessibility and burial

Solvent accessible surface area was calculated using the algorithm of Lee & Richards (1971) as implemented in the program *access* by Simon Hubbard (unpublished). We used the probe *contact* surface area in preference to the probe centre *locus* surface, and a probe of radius 1.4 Å. Accessible contact surface areas (*a.c.s.a.*) are smaller than probe locus surfaces and molecular surfaces by a factor of approximately 3.4 and 2.8, respectively.

An accessible residue is defined as having an *a.c.s.a.* > 1.0 Å^2. An antigen contacting residue is defined by burial from solvent by antigen binding such that:

$$\Delta_c^u a.c.s.a. > 1.0 \text{ Å}^2$$

where

$$\Delta_c^u a.c.s.a. = a.c.s.a._{uncomplexed} - a.c.s.a._{complexed}$$

Fraction burial per residue for residue r, ρ_r, is defined as the fraction of solvent accessible surface area buried by antigen:

$$\rho_r = \frac{\Delta_c^u a.c.s.a.}{a.c.s.a._{uncomplexed}}$$

Mean fraction burial, $\tilde{\rho}$, at residue r is defined as:

$$\tilde{\rho}_r = \frac{1}{N} \sum_{a=1}^{N} \rho_r$$

where N is the number of antibody structures possessing residue r with *a.c.s.a.* > 0.

Canonical loops and structural alignment

We have used an automated loop clustering program (Martin & Thornton, 1996) to assign each CDR from our dataset to one of a number of distinct backbone conformations. The first step in the protocol employs Ward's minimum variance method (Ward, 1963; Wishart, 1969) to cluster "loop vectors" derived from backbone torsion angles. A second "post-clustering" step merges clusters which are similar in cartesian space. In most cases the conformational clusters match the canonical classes described by Chothia *et al.* (1989), but additional conserved backbone conformations have also been identified.

CDRs were aligned for the "through-space" contact analysis using the program *ProFit*†, fitting on three framework residues either side of the loop residues (see Table 3). In the Figures, single representatives for multiply populated CDR conformations were chosen alphabetically by PDB code.

Surface analysis

In our surface topography analysis we have used the fractal atomic density measure of Kuhn *et al.* (1992) to determine local curvature on a molecular surface. This quantity is calculated for each surface point as the slope of the log-log plot (by linear regression) of the number of (all) protein atom centres contained within a 1 Å thick spherical shell *versus* the internal radius of this shell as it varies between 0 and 10 Å. The fractal atomic density of a point on a perfectly flat protein surface would be 2.0 because the number of atom centres within each shell would be proportional to the sphere's surface area (which is proportional to the radius raised to the power of 2.0). Convex and concave surface points give fractal atomic densities less than and greater than 2.0, respectively.

† http://www.biochem.ucl.ac.uk/~martin/#programs.

Acknowledgements

R.M.M. thanks the UK Medical Research Council and Glaxo Wellcome for support. A.C.R.M. thanks the UK Medical Research Council for support.

References

Arevalo, J. H., Hassig, C. A., Stura, E. A., Sims, M. J., Taussig, M. J. & Wilson, I. A. (1994). Structural-analysis of antibody specificity - detailed comparison of 5 Fab'-steroid complexes. *J. Mol. Biol.* **241**, 663–690.

Bernstein, F. C., Koetzle, T. F., Williams, G. J. B., Meyer, E. F., Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T. & Tasumi, M. (1977). The protein data bank: A computer-based archival file for macromolecular structures. *J. Mol. Biol.* **112**, 535–542.

Brinkman, K., Termaat, R., Berden, J. H. M. & Smeenk, R. J. T. (1990). Anti-DNA antibodies and lupus nepritis: the complexity of cross-reactivity. *Immun. Today*, **11**, 232–234.

Brünger, A. T., Leahy, D. J., Hynes, T. R. & Fox, R. O. (1991). 2.9 Å-resolution structure of an anti-dinitrophenyl-spin-label monoclonal-antibody Fab fragment with bound hapten. *J. Mol. Biol.* **221**, 239–256.

Chester, K. A. & Hawkins, R. E. (1995). Clinical issues in antibody design. *Trends Biotechnol.* **13**, 294–300.

Chothia, C. & Lesk, A. M. (1987). Canonical structures for the hypervariable regions of immunoglobulins. *J. Mol. Biol.* **196**, 901–917.

Chothia, C., Lesk, A. M., Levitt, M., Amit, A. G., Mariuzza, R. A., Phillips, S. E. V. & Poljak, R. J. (1986). The predicted structure of immunoglobulin D1.3 and its comparison with the crystal structure. *Science*, **233**, 755–758.

Chothia, C., Lesk, A. M., Tramontano, A., Levitt, M., Smith-Gill, S. J., Air, G., Sheriff, S., Padlan, E. A., Davies, D. R., Tulip, W. R., Colman, P. M., Alzri, P. M. & Poljak, R. J. (1989). Conformations of immunoglobulin hypervariable regions. *Nature*, **342**, 877–883.

Churchill, M. E. A., Stura, E. A., Pinilla, C., Appel, J. R., Houghten, R. A., Kono, D. H., Balderas, R. S., Fieser, G. G., Schulzegahmen, U. & Wilson, I. A. (1994). Crystal-structure of a peptide complex of antiinfluenza peptide antibody-Fab-26/9 - comparison of two different antibodies bound to the same peptide antigen. *J. Mol. Biol.* **241**, 534–556.

Delaney, J. S. (1992). Finding and filling protein cavities using cellular logic operations. *J. Mol. Graph.* **101**, 174.

Gerstein, M. (1992). A resolution-sensitive procedure for comparing protein surfaces and its application to the comparison of antigen combining sites. *Acta Crystallog. sect. A*, **48**, 271–276.

Herron, J. N., He, X. M., Ballard, D. W., Blier, P. R., Pace, P. E., Bothwell, A. L. M., Voss, E. W. & Edmundson,

A. B. (1991). An autoantibody to single-stranded-DNA - comparison of the 3-dimensional structures of the unliganded Fab and a deoxynucleotide Fab complex. *Proteins: Struct. Funct. Genet.* **11**, 159–175.

Hilvert, D. (1994). Catalytic antibodies. *Curr. Opin. Struct. Biol.* **4**, 612–617.

Jeffrey, P. D., Schildbach, J. F., Chang, C. Y. Y., Kussie, P. H., Margolies, M. N. & Sheriff, S. (1995). Structure and specificity of the antidigoxin antibody-40-50. *J. Mol. Biol.* **248**, 344–360.

Killard, A. J., Deasy, B., O'Kennedy, R. & Smyth, M. R. (1995). Antibodies - production, functions and applications in biosensors. *Trends Anal. Chem.* **14**, 257–266.

Kuhn, L. A., Siani, M. A., Pique, M. E., Fisher, C. L., Getzoff, E. D. & Tainer, J. A. (1992). The interdependence of protein surface topography and bound water molecules revealed by surface accessibility and fractal density measures. *J. Mol. Biol.* **228**, 13–22.

Lawrence, M. C. & Colman, P. M. (1993). Shape complementarity at protein/protein interfaces. *J. Mol. Biol.* **234**, 946–950.

Lee, B. K. & Richards, F. M. (1971). The interpretation of protein structures: Estimation of static accessibility. *J. Mol. Biol.* **55**, 379–400.

Lescar, J., Pellegrini, M., Souchon, H., Tello, D., Poljak, R. J., Peterson, N., Greene, M. & Alzari, P. M. (1995). Crystal-structure of a cross-reaction complex between Fab F9.13.7 and guinea-fowl lysozyme. *J. Biol. Chem.* **270**, 18067–18076.

Love, R. A., Villafranca, J. E., Aust, R. M., Nakamura, K. K., Jue, R. A., Major, J. G., Radhakrishnan, R. & Butler, W. F. (1993). How the anti-(metal chelate) antibody CHA255 is specific for the metal-ion of its antigen - X-ray structures for 2 Fab' hapten complexes with different metals in the chelate. *Biochemistry*, **32**, 10950–10959.

Martin, A. C. R. & Thornton, J. M. (1996). Structural families in loops of homologous proteins: Automatic classification, modelling and application to antibodies. *J. Mol. Biol.* in the press.

Martin, A. C. R., Cheetham, J. C. & Rees, A. R. (1991). Molecular modelling of antibody combining sites. *Meth. Enzymol.* **203**, 121–153.

Nicholls, A., Sharp, K. A. & Honig, B. (1991). Protein folding and association - insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins: Struct. Funct. Genet.* **11**, 281–296.

Padlan, E. A., Silverton, E. W., Sheriff, S., Cohen, G. H., Smithgill, S. J. & Davies, D. R. (1989). Structure of an antibody antigen complex - crystal structure of the HyHEL-10 Fab-lysozyme complex. *Proc. Natl Acad. Sci. USA*, **86**, 5938–5942.

Padlan, E. A., Abergel, C. & Tipper, J. P. (1995). Identification of specificity-determining residues in antibodies. *FASEB J.* **9**, 133–139.

Pokkuluri, P. R., Bouthillier, F., Li, Y. G., Kuderova, A., Lee, J. & Cygler, M. (1994). Preparation, characterization and crystallization of an antibody Fab fragment that recognizes RNA:crystal-structures of native Fab and three Fab-mononucleotide complexes. *J. Mol. Biol.* **243**, 283–297.

Prasad, L., Vandonselaar, M., Lee, J. S. & Delbaere, L. T. J. (1988). Structure determination of a monoclonal Fab fragment specific for histidine-containing protein of the phosphoenolpyruvate - sugar phosphotransferase system of *Escherichia coli*. *J. Biol. Chem.* **263**, 2571–2574.

Rees, A. R., Staunton, D., Webster, D. M., Searle, S. J., Henry, A. H. & Pedersen, J. T. (1994). Antibody design: beyond the natural limits. *Trends Biotechnol.* **12**, 199–206.

Rini, J. M., Stanfield, R. L., Stura, E. A., Salinas, P. A., Profy, A. T. & Wilson, I. A. (1993). Crystal-structure of a human-immunodeficiency-virus type-1 neutralizing antibody, 50.1, in complex with its V3 loop peptide antigen. *Proc. Natl Acad. Sci. USA*, **90**, 6325–6329.

Satow, Y., Cohen, G. H., Padlan, E. A. & Davies, D. R. (1986). Phosphocholine binding immunoglobulin Fab McPC603 an X-ray-diffraction study at 2.7 Å. *J. Mol. Biol.* **190**, 593–604.

Schulzegahmen, U., Rini, J. M. & Wilson, I. A. (1993). Detailed analysis of the free and bound conformations of an antibody: X-ray structures of Fab 17/9 and three different Fab-peptide complexes. *J. Mol. Biol.* **234**, 1098–1118.

Stanfield, R. L., Takimotokamimura, M., Rini, J. M., Profy, A. T. & Wilson, I. A. (1993). Major antigen-induced domain rearrangements in an antibody. *Structure*, **1**, 83–93.

Tulip, W. R., Varghese, J. N., Laver, W. G., Webster, R. G. & Colman, P. M. (1992). Refined crystal-structure of the influenza virus-N9 neuraminidase NC41 Fab complex. *J. Mol. Biol.* **227**, 122–148.

Ward, J. H. (1963). Hierachical grouping to optimize an objective function. *J. Am. Statist. Assoc.* **58**, 236–244.

Webster, D. M., Henry, A. H. & Rees, A. R. (1994). Antibody-antigen interactions. *Curr. Opin. Struct. Biol.* **4**, 123–129.

Wilson, I. A. & Stanfield, R. L. (1993). Antibody-antigen interactions. *Curr. Opin. Struct. Biol.* **3**, 113–118.

Wilson, I. A. & Stanfield, R. L. (1994). Antibody-antigen interactions - new structures and new conformational-changes. *Curr. Opin. Struct. Biol.* **4**, 857–867.

Winter, G., Griffiths, A. D., Hawkins, R. E. & Hoogenboom, H. R. (1994). Making antibodies by phage display technology. *Annu. Rev. Immunol.* **12**, 433–455.

Wishart, D. (1969). An algorithm for hierachical classifications. *Biometrics*, **25**, 165–170.

Wong, Y. W., Kussie, P. H., Parhamiseren, B. & Margolies, M. N. (1995). Modulation of antibody-affinity by an engineered amino-acid substitution. *J. Immunol.* **154**, 3351–3358.

Wu, T. T. & Kabat, E. A. (1970). An analysis of the sequences of the variable regions of Bence Jones proteins and myeloma light chains and their implications for antibody complementarity. *J. Exp. Med.* **132**, 211–250.

Edited by F. E. Cohen

(Received 19 December 1995; received in revised form 23 July 1996; accepted 30 July 1996)

Pergamon

0161-5890(93)E0005-7

Molecular Immunology, Vol. 31, No. 3, pp. 169–217, 1994
Elsevier Science Ltd. Printed in Great Britain
0161-5890/94 $6.00 + 0.00

REVIEW

ANATOMY OF THE ANTIBODY MOLECULE

EDUARDO A. PADLAN

Laboratory of Molecular Biology, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, MD 20892, U.S.A.

(*Received* 1 *June* 1993; *accepted* 1 *October* 1993)

Abstract—The structures of the various regions of an antibody molecule are analysed and correlated with biological function. The structural features which relate to potential applications are detailed.

Key words: X-ray structures, antibody–ligand interactions, polyreactivity, humanization strategies.

INTRODUCTION

In terms of structure, antibodies are probably now the most studied of all proteins. Amino acid and nucleotide sequence data are available for thousands of different chains (Kabat *et al.*, 1991) and three-dimensional structures, obtained by X-ray crystallographic analysis, are available for whole antibodies and for a variety of fragments. Very recently, all parts of an antibody molecule have been visualized (Harris *et al.*, 1992), including the hinge which, previously, either was only partly discernible in the electron-density map because of its flexibility (Marquart *et al.*, 1980), or was deleted (Sarma *et al.*, 1971; Silverton *et al.*, 1977; Sarma and Laudin, 1982; Rajan *et al.*, 1983) in the intact antibodies analysed by X-ray diffraction.

The antibody fragments whose structures have been elucidated crystallographically include several human and murine V_L dimers (Epp *et al.*, 1974, 1975; Fehlhammer *et al.*, 1975; Colman *et al.*, 1977; Furey *et al.*, 1983; Stevens *et al.*, 1981; Steipe *et al.*, 1992), Fv from murine and human antibodies (Bhat *et al.*, 1990; Fan *et al.*, 1992) including a "humanized" murine Fv (Eigenbrot *et al.*, 1993), several Bence–Jones light-chain dimers (Schiffer *et al.*, 1973, 1989; Ely *et al.*, 1985), Fc from human and rabbit IgG (Deisenhofer, 1981; Sutton and Phillips, 1983) and pFc′ from guinea pig IgG (Bryant *et al.*, 1985). The structure of the complex of human IgG1 Fc and fragment B of protein A from *Staphylococcus aureus* has also been determined (Deisenhofer, 1981).

By far, the fragment which has been the most studied is the Fab, with the structures of several dozen Fabs of different ligand-binding specificities having been elucidated by X-ray diffraction, many with bound specific ligand. The ligands range in size from small haptens to macromolecules. The small ligands include phosphocholine (Padlan *et al.*, 1973, 1985; Segal *et al.*, 1974), a hydroxy derivative of vit. K_1 (Amzel *et al.*, 1974), a dinitrophenyl-spin-label hapten (Bruenger *et al.*, 1991), fluorescein (Herron *et al.*, 1989), digoxin (Bruenger, 1991) and progesterone (Wilson *et al.*, 1991). The bigger ligands include oligopeptides, like cyclosporin A (Altschuh *et al.*, 1992; Vix *et al.*, 1993), angiotensin II (Garcia *et al.*, 1992), a 19-amino acid peptide homolog of a myohemerythrin helix (Stanfield *et al.*, 1990), a 15-residue long peptide derived from cholera toxin (Shoham *et al.*, 1991) and a nonapeptide from influenza virus hemagglutinin (Rini *et al.*, 1992), and fragments of other macromolecular antigens, like a trinucleotide of deoxythymidylic acid (Herron *et al.*, 1991) and a trisaccharide epitope of a branched bacterial lipopolysaccharide (Cygler *et al.*, 1991). The macromolecular ligands include lysozyme (Amit *et al.*, 1986; Sheriff *et al.*, 1987; Padlan *et al.*, 1989; Fischmann *et al.*, 1991; Lescar *et al.*, 1993), influenza virus neuraminidase (Colman *et al.*, 1987, 1989; Tulip *et al.*, 1989, 1992*a*, *b*), HIV-1 reverse transcriptase (Arnold *et al.*, 1992), and even another Fab in an idiotope–anti-idiotope complex (Bentley *et al.*, 1990). In addition, X-ray structures are available for an Fab complexed with streptococcal protein G (Derrick and Wigley, 1992), for a chimeric Fab in which the variable domains were from a murine antibody and the constant domains were from human molecules (Brady *et al.*, 1992), and for two versions of a "humanized" murine Fab (Eigenbrot *et al.*, 1993).

The availability, in some cases, of both complexed and uncomplexed structures for the same antibody permits the evaluation of the possibility of conformational changes occurring upon ligand binding. Previously, with phosphocholine and vit. K_1OH, no conformational changes in the Fabs were observed when these small ligands were bound in the crystal (Padlan *et al.*, 1973, 1985; Segal *et al.*, 1974; Amzel *et al.*, 1974). More recent studies with larger ligands reveal that significant changes can occur upon complexation (Bhat *et al.*, 1990; Stanfield *et al.*, 1990; Herron *et al.*, 1991; Wilson *et al.*, 1991; Rini *et al.*, 1992).

Starting soon after crystallographic data became available, various aspects of antibody structure and function have been the subject of analysis, including the distribution of the different amino acid types in relation to ligand-binding specificity (e.g. Kabat *et al.*, 1977;

Best Available Copy

Table 1. Immunoglobulin structures which have been determined by X-ray crystallography

Antibody	Isotype	Fragment	PDB Code	Resolution	R-value	Ligand	Reference
(Human)							
Dob	(IgG1,κ)	whole Ig		4.0			Sarma and Laudin (1982)
Kol	(IgG1,λ)	whole Ig	2IG2	3.0	0.207		Marquart *et al.* (1980)
Mcg	(IgG1,λ)	whole Ig		2.8			[a]
NEW	(IgG1,λ)	Fab	7FAB	2.0	0.169		Saul and Poljak (1992)
		Fab		3.5		vit. K_1OH	Amzel *et al.* (1974)
Kol	(IgG1,λ)	Fab	2FB4	1.9	0.189		Marquart *et al.* (1980)
Hil	(IgG1,λ)	Fab	8FAB	1.8	0.173		[b]
3D6	(IgG1,κ)	Fab	1DFB	2.7	0.177		He *et al.* (1992)
POT	(IgM,κ)	Fv	1IGM	2.3	0.201		Fan *et al.* (1992)
	(IgG1)	Fc	1FC1	2.9	0.22		Deisenhofer (1981)
	(IgG1)	Fc	1FC2	2.8	0.24	fragment B of Protein A	Deisenhofer (1981)
Mcg	(λ)	L-dimer	2MCG	2.0	0.187		Ely *et al.* (1989)
			3MCG	2.0	0.208		Ely *et al.* (1989)
LOC	(λ)	L-dimer	1BJL	3.0	0.194		Chang *et al.* (1985)
		L-dimer	2BJL	2.8	0.216		Schiffer *et al.* (1989)
Mcg-Weir	(λ)	L-dimer	1MCW	3.5	0.170		Ely *et al.* (1985)
REI	(κ)	V_L-dimer	1REI	2.0	0.24		Epp *et al.* (1975)
Au	(κ)	V_L-dimer		2.5	0.31		Fehlhammer *et al.* (1975)
Rhe	(λ)	V_L-dimer	2RHE	1.6	0.149		Furey *et al.* (1983)
ROY	(κ)	V_L-dimer		3.0	0.33		Colman *et al.* (1977)
Wat	(κ)	V_L-dimer					Stevens *et al.* (1981)
("Humanized" murine)							
4D5		Fv	1FVC	2.2	0.183		Eigenbrot *et al.* (1993)
		Fab	1FVD	2.5	0.179		Eigenbrot *et al.* (1993)
		Fab	1FVE	2.7	0.171		Eigenbrot *et al.* (1993)
(Murine–human chimera)							
B72.3		Fab	1BBJ	3.1	0.176		Brady *et al.* (1992)
(Murine)							
Mab231	(IgG2a)	whole Ig		3.5	0.188		Harris *et al.* (1992)
McPC603	(IgA,κ)	Fab	1MCP	2.7	0.225		Satow *et al.* (1986)
		Fab		3.1	0.196	phosphocholine	[c]
J539	(IgA,κ)	Fab	2FBJ	1.95	0.194		[d]
D1.3	(IgG1,κ)	Fab	1FDL	2.5	0.184	lysozyme	Fischmann *et al.* (1991)
		Fab		2.6	0.27		Bentley *et al.* (1989)
NC41	(IgG2a,κ)	Fab	1NCA	2.5	0.191	influenza virus neuraminidase	Tulip *et al.* (1992a)
		Fab	1NCC	2.5	0.212	neuraminidase I368R mutant	Tulip *et al.* (1992b)
		Fab	1NCB	2.5	0.165	neuraminidase N329D mutant	Tulip *et al.* (1992b)
		Fab	1NCD	2.9	0.157	influenza virus neuraminidase	Tulip *et al.* (1992a)
HyHEL-5	(IgG1,κ)	Fab	2HFL	2.54	0.245	lysozyme	Sheriff *et al.* (1987)
HED10		Fab		3.0	0.272		Cygler *et al.* (1987)
CF4C4	(IgG1,κ)	Fab		4.0	0.470		Vitali *et al.* (1987)
Jel42		Fab		3.5	0.282		Prasad *et al.* (1988)
NC10		Fab		3.0	0.20	influenza virus neuraminidase	Tulip *et al.* (1989)
HyHEL-10	(IgG1,κ)	Fab	3HFM	3.0	0.246	lysozyme	Padlan *et al.* (1989)
4-4-20	(IgG2a,κ)	Fab	4FAB	2.7	0.215	fluorescein	Herron *et al.* (1989)
R19.9	(IgG2b,κ)	Fab	2F19	2.8	0.182		Lascombe *et al.* (1992)
		Fab	1FAI	2.7	0.189		Lascombe *et al.* (1992)
E225	(IgG2b,κ)	Fab		2.5	0.179	D1.3 Fab	Bentley *et al.* (1990)
B1312	(IgG1,κ)	Fab	2IGF	2.8	0.22	myohemerythrin peptide	Stanfield *et al.* (1990)
		Fab	1IGF	2.8	0.18		Stanfield *et al.* (1990)
MAb131	(IgG1,κ)	Fab		3.0	0.25	angiotensin II	Garcia *et al.* (1991)
BV04-01	(IgG2b,κ)	Fab		2.66	0.191	$d(pT)_3$	Herron *et al.* (1991)
		Fab		2.0	0.245		Herron *et al.* (1991)

Table 1. Continued overleaf.

Table 1.—*Continued*

36-71	(IgG1,κ)	Fab	6FAB	1.85	0.248		Strong *et al.* (1991)
AN02	(IgG1,κ)	Fab	1BAF	2.9	0.195	DNP-spin-label hapten	Bruenger *et al.* (1991)
NQ10/12.5	(IgG1,κ)	Fab		2.8	0.182		Alzari *et al.* (1990)
		Fab		3.0	0.19	2-phenyl-oxazolone	Alzari *et al.* (1990)
TE33		Fab		3.0	0.386	cholera toxin peptide	Shoham *et al.* (1991)
Se155-4	(IgG1,λ)	Fab		2.05	0.185	dodecasaccharide	Cygler *et al.* (1991)
26-10		Fab		2.7	0.17	digoxin	Bruenger (1991)
		mutant Fab		2.5	0.21		Bruenger (1991)
B13A2		Fab					Wilson *et al.* (1991)
HC19	(IgG1,λ)	Fab		3.5	0.176		Bizebard *et al.* (1991)
DB3	(IgG1,κ)	Fab		2.8			Wilson *et al.* (1991)
		Fab		2.4		progesterone	Wilson *et al.* (1991)
		Fab		2.8		iodobenzoyl progesterone	Wilson *et al.* (1991)
		Fab		3.0		11α-hemisuccinyl progesterone	Wilson *et al.* (1991)
R454511	(IgG1,κ)	Fab		2.65	0.185	cyclosporin A	Altschuh *et al.* (1992)
mAb28		Fab		7		HIV-1 reverse transcriptase and DNA	Arnold *et al.* (1992)
Jel72		Fab					[e]
Jel318		Fab					[f]
		Fab				protein G	Derrick and Wigley (1992)
Yst9-1	(IgG2b,κ)	Fab	1MAM	2.5	0.215		[g]
17/9	(IgG2a,κ)	Fab	1HIN	3.1	0.22	hemagglutinin peptide	Rini *et al.* (1992)
		Fab	1HIM	2.9	0.20	hemagglutinin peptide	Rini *et al.* (1992)
		Fab	1HIL	2.0	0.19		Rini *et al.* (1992)
8F5	(IgG2a,κ)	Fab	1BBD	2.8	0.190		Tormo *et al.* (1992)
F9.13.7		Fab		4.0	0.409	guinea-fowl lysozyme	Lescar *et al.* (1993)
D1.3		Fv		2.4	0.19	lysozyme	Bhat *et al.* (1990)
		Fv		1.9	0.19		
McPC603		V_L	1IMM	2.00	0.149		Steipe *et al.* (1992)
		V_L,CDR-1 replaced	1IMN	1.97	0.149		Steipe *et al.* (1992)
(Rabbit)							
	(IgG)	Fc		2.7			Sutton and Phillips, (1983)
(Guinea pig)							
	(IgG1)	pFc'	1PFC	3.125	0.303		Bryant *et al.* (1985)

[a] Guddat, Edmundson and Herron (to be published), cited in Kabat *et al.* (1991).
[b] Saul and Poljak (to be published), cited in PDB Entry: 8FAB.
[c] Padlan, Cohen and Davies (in preparation).
[d] Bhat, Padlan and Davies (in preparation).
[e] Mol, Muir, Lee and Anderson (unpublished), cited in Kabat *et al.* (1991).
[f] Muir, Cygler, Mol, Lee and Anderson (unpublished), cited in Kabat *et al.* (1991).
[g] Rose, Przybylska, To, Kayden, Oomen, Vorberg, Young and Bundle (to be published). cited in PDB Entry: 1MAM.

Padlan, 1990*a*; Mian *et al.*, 1991), interdomain interactions (e.g. Poljak *et al.*, 1975*a*, *b*, 1976; Davies *et al.*, 1975*a*, *b*; Amzel and Poljak, 1979; Davies and Metzger, 1983; Novotny *et al.*, 1983; Novotny and Haber, 1985; Chothia *et al.*, 1986; Padlan *et al.*, 1986; Schiffer *et al.*, 1988; Miller, 1990), flexibility (e.g. Edmundson *et al.*, 1974, 1978, 1987; Burton, 1985, 1990*a*, *b*; Huber and Bennett, 1987; Lesk and Chothia, 1988; Ely *et al.*, 1989; Rini *et al.*, 1992; Davies and Padlan, 1992), hypervariable region structures (e.g. Padlan and Davies, 1975; Potter *et al.*, 1976; Padlan, 1977*a*; De la Paz *et al.*, 1986; Chothia and Lesk, 1987; Chothia *et al.*, 1989, 1992; Tramontano *et al.*, 1990; Wu *et al.*, 1993), disulfide bridges (e.g. Steiner, 1985), and antibody–ligand interactions (e.g. Padlan *et al.*, 1976; Huber, 1986; Mariuzza *et al.*, 1987; Alzari *et al.*, 1988; Colman, 1988; Davies *et al.*, 1988, 1990; Wilson *et al.*, 1991; Davies and Chacko, 1993). In addition to X-ray crystallography, other techniques have been used to study antibody structure, including two-dimensional NMR (e.g. Anglister, 1990; Theriault *et al.*, 1991), genetic engineering (e.g. Morrison *et al.*, 1984; Morrison and Oi,

Best Available Copy

1988; Bird *et al.*, 1988; Huston *et al.*, 1988; Winter, 1989; Tan *et al.*, 1990; Helm *et al.*, 1991; Glockshuber *et al.*, 1992; Jin *et al.*, 1992) and theoretical analysis (e.g. Novotny *et al.*, 1989; Novotny and Sharp, 1992).

Here, we will analyse the currently available three-dimensional data on antibodies which have been obtained by crystallographic analysis, focusing on those structural features which relate to antibody function and potential applications. The crystallographically-determined structures, which are known to the author, are listed in Table 1. The present analysis is limited to those structures for which atomic coordinates are available at the time of writing from the Protein Data Bank (PDB) (Bernstein *et al.*, 1977; Abola *et al.*, 1987), or have been kindly made available by the original investigators.

The studies on antibody structure are too many to enumerate and only a limited number of references could reasonably be cited. It is hoped that the citations given can serve as starting points for those who are interested in particular aspects of antibody structure. Several comprehensive reviews have been written recently on the subject (e.g. Marguzza *et al.*, 1987; Alzari *et al.*, 1988; Colman, 1988; Davies *et al.*, 1988, 1990; Wilson *et al.*, 1991; Davies and Chacko, 1993) and the reader is encouraged to consult those also.

Caveat

Only a few of the structures listed in Table 1 have been determined to high resolution (2.0 Å or better) and refined extensively (to crystallographic R-values of 0.20 or better), so that the errors associated with many of the structures could be large. An example of a highly-refined structure is J539 Fab, which has been determined to 1.95-Å resolution and refined to an R-value of 0.194 (PDB Entry: 2FBJ); for J539 Fab, the error in the atomic coordinates, estimated by the method of Luzzati (1953), is 0.25 Å (Bhat, Padlan and Davies, unpublished results). An example of a structure determined at medium resolution is HyHEL-5 Fab, which has been determined to 2.54-Å resolution and refined to an R-value of 0.245 (Sheriff *et al.*, 1987); the estimated error for HyHEL-5 Fab is 0.40 Å. These error estimates are for the most ordered (usually, interior) regions of the molecule; the errors will be larger for exposed segments, especially for the side chain atoms. For structures that have been determined to lower resolution or that have been refined less extensively, the errors are probably higher. In view of the potentially large errors in the individual structures, it would be prudent to look for trends from the analysis of the whole set of structures than to draw conclusions on the basis of only one.

THE STRUCTURE OF ANTIBODIES

An antibody molecule is composed of three major fragments: the two Fabs, which are identical and each of which contains the light chain and the first two domains of the heavy chain, and the Fc, which contains the C-terminal constant domains of the two heavy chains. The Fabs are linked to the Fc by the hinge region, which varies in length and flexibility in the different antibody classes and isotypes. The antigen-binding sites (paratopes) are located at the tips of the Fabs. A representation of human IgG1 antibody is shown in Fig. 1.

From the earliest studies of antibody structure (Poljak *et al.*, 1973, 1974; Schiffer *et al.*, 1973; Epp *et al.*, 1974; Segal *et al.*, 1974), it was clear that all antibody domains, whether variable or constant, form compact globular structures with a characteristic fold, termed the *Immunoglobulin Fold* by Poljak *et al.* (1972). Each domain consists of a stable arrangement of hydrogen-bonded, anti-parallel β-strands which form a bilayer structure, further stabilized by a disulfide bond between the two layers. In the variable domains, the bilayer structure is formed by nine strands; in the constant domains, the bilayer is formed by seven strands. Bends of different sizes and conformations connect the strands. The predominant secondary structure in antibodies is the anti-parallel β-sheet, with short stretches of α-helix found in some bends.

The variable domains of the light and heavy chains, the V_L and V_H, are similar to each other in three-dimensional structure, as are the constant domains: the C_L of the light chain and the C_H1 of the heavy chain in the Fab, and the C_H2 and the C_H3 domains in Fcγ. Homologous domains from different species are very similar and are essentially superposable. Even in the variable domains, where as much as 30% of the residues could be different among different antibodies, the differences in structure are almost entirely confined to the hypervariable regions (Wu and Kabat, 1970), and usually only if length differences exist in those regions (Padlan and Davies, 1975).

The availability of three-dimensional data for a variety of fragments from different antibody classes and isotypes and from different species, helps in the identification of structurally important features and of analogous structural elements. An attempt to align representative sequences from the various human antibody chain types on the basis of structure is made in Table 2. For purposes of this review, the C-terminus of the Fd, the heavy-chain component of the Fab, is defined as the Cys residue at position 220 [Eu numbering (Edelman *et al.*, 1969)] in human IgG1, or the analogous residue in the other heavy chain classes; the disulfide bridge, which Cys220 forms with the Cys at 214 in the light chain, creates a natural (structural) boundary for the Fab. Accordingly, the hinge is defined as starting with the Asp221 in human IgG1, or analogous residue. In the only available Fc structure (PDB Entries: 1FC1 and 1FC2), the first visible residue, which is probably the first residue in the compact portion of the Fc, is Pro238. Again, for purposes of this review only, we will define the Fc as starting with the Pro238 in human IgG1, or analogous residue in the other heavy chain classes, and define the end of the hinge as the residue preceding Pro238.

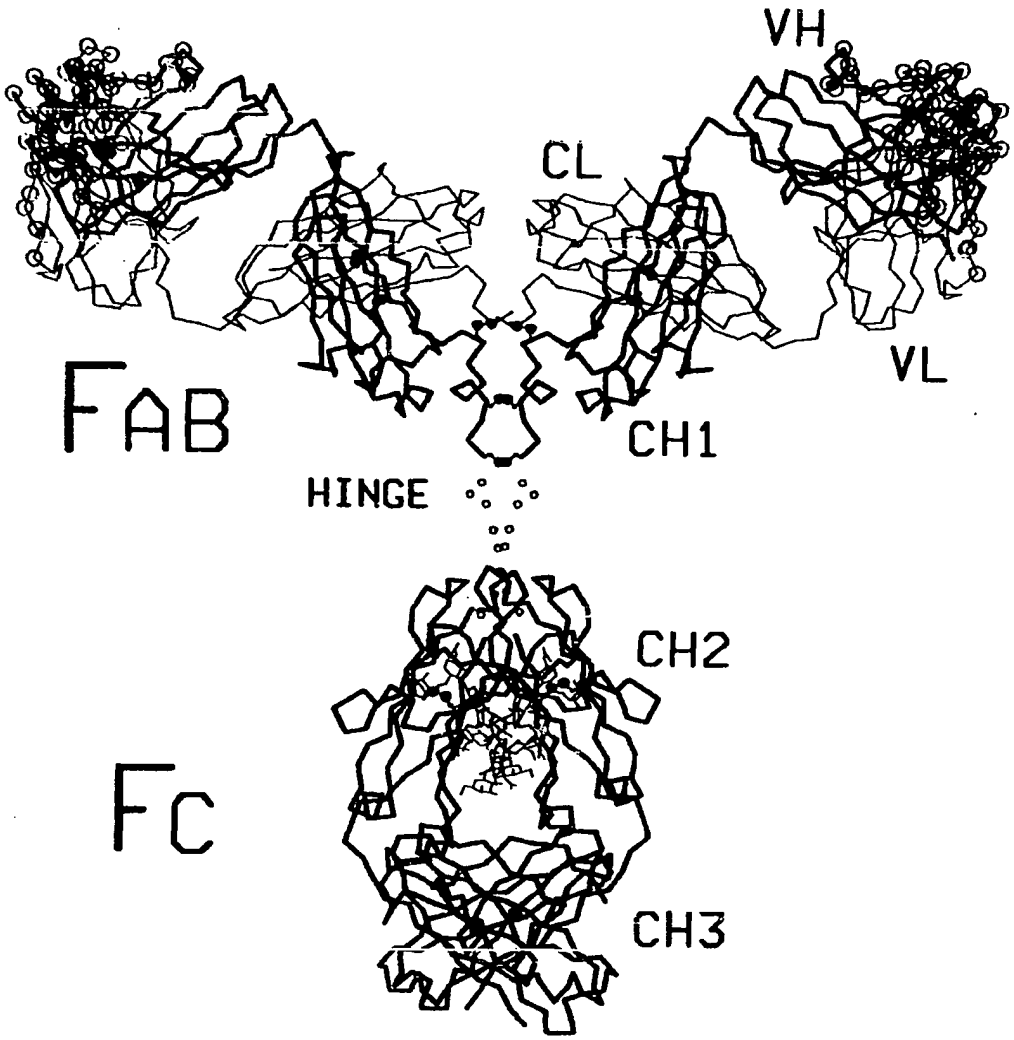


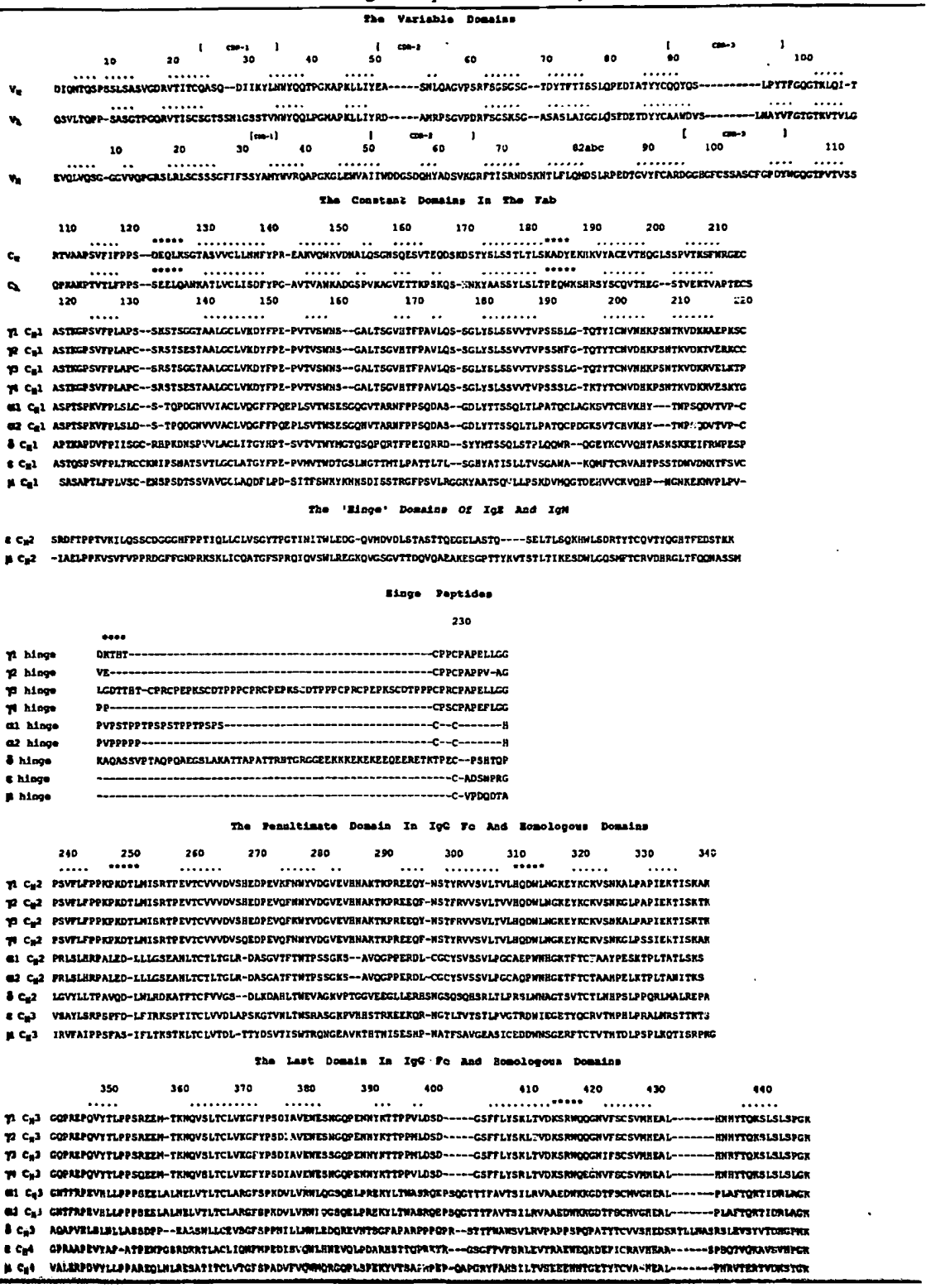
Fig. 1. Line drawing representation of a possible structure for human IgG1. The drawing is a composite of the crystal structures of human IgG1, Kol [Marquart *et al.* (1980); PDB Entry: 21G2], which is visible only to the inter-heavy chain cystine at 229 (Eu numbering), and of human IgG1 Fc [Deisenhofer (1981); PDB Entry: 1FC2], which is visible only from Pro238 to Leu443 (Eu numbering). The octapeptide, PAPELLGG, between Cys229 and Pro238 has not been visualized and its probable location is indicated by the unconnected circles in the middle of the figure. Only the α-carbons are traced. The heavy chains are drawn with thick lines, the light chains with thin lines. The domains of the light chain (V_L and C_L) and those of the heavy chain (V_H, C_H1, C_H2 and C_H3) are labeled. Also indicated is the hinge, the region that lies between the Fabs and the Fc and which varies considerably in size in the different antibody types. The CDR residues are drawn with larger open circles at the tips of the Fabs. The disulfide bridges are indicated by filled circles. The carbohydrate moieties attached to the asparagines at position 297 (Eu numbering) are drawn with thin lines between the two C_H2 domains.

The structure of Fabs

The Fab structures which are examined here are those of the murine J539 (PDB Entry: 2FBJ), McPC603 with and without bound phosphocholine (PDB Entry: 1MCP and unpublished results of Padlan, Cohen and Davies), HyHEL-10 (PDB Entry: 3HFM), HyHEL-5 (PDB Entry: 2HFL), R19.9 (PDB Entry: 2F19), 4-4-20 (PDB Entry: 4FAB), BV04-01 with and without bound d(pT)$_2$ (atomic coordinates kindly made available by Dr Allen B. Edmundson and coworkers), 36-71 (PDB Entry: 6FAB), B13I2 with and without bound peptide (PDB Entries: 2IGF and 1IGF, respectively), D1.3 (PDB Entry: 1FDL), Yst9-1 (PDB Entry: 1MAM), AN02 (PDB Entry: 1BAF), 17/9 with ligand in two crystal forms (PDB Entry: 1HIN and 1HIM, the latter with two Fabs in the asymmetric unit of the crystal) and without ligand (PDB Entry: 1HIL, with two Fabs in the asymmetric unit of the crystal), NC41 with bound neuraminidase (PDB Entry: 1NCA) and 8F5 (PDB Entry: 1BBD), and those of the human Kol (PDB Entry: 2FB4), NEW (PDB Entry: 7FAB), Hil (PDB Entry: 8FAB, with two

Best Available Copy

Table 2. Structural alignment of human antibody domains



Secondary structure from program DSSP of Kabsch and Sander (1983) on 3D6 (PDB Entry: 1DFB) for Vκ and Cκ, on Kol (PDB Entry: 2FB4) for Vλ, Cλ, V_H and C_H1, and on human IgG1 Fc (PDB Entry: 1FC1) for C_H2 and C_H3. Residues which participate in the β-sheets are indicated by (.); those which are in α-helices are indicated by (*). The numbering scheme for the V_L, V_H and C_L domains follows the convention of Kabat *et al.* (1991); Eu numbering (Edelman *et al.*, 1969) is used for the IgG1 constant domains. The domains of J539 and Kol Fab, which are two of the most highly-refined Fab structures listed in Table 1, were used in the structural alignment; the other sequences were aligned on the basis of their sequence similarity to these two.

Fabs in the asymmetric unit of the crystal) and 3D6 (PDB Entry: 1DFB). The human Fv, POT (PDB Entry: 1IGM), was included in the analysis where appropriate.

In the Fab (Fig. 2), V_L and V_H associate closely to form a compact module, the Fv, and are related by a pseudo-dyad; the constant domains in the Fab, the C_L and the C_H1, likewise form a compact module and are also related by a pseudo-dyad (Table 3). On average, the V_L–V_H contact buries approximately 1420 Å^2 of surface area (720 in V_L and 700 in V_H). The V_L–V_H contact varies from antibody to antibody (Table 4) and differences are found even for the same Fab but crystallized in different crystal forms. The C_L–C_H1 contact (Table 4) buries approximately 1710 Å^2 of surface area (870 in C_L and 840 in C_H1) and, here also, some variation is found in the contacts present in supposedly identical C_L–C_H1 pairings. The variation seen in the identical C_L–C_H1 pairings may reflect the different lattice forces to which the molecules are subjected in the different crystal forms, the errors in the crystallographic analysis of these structures, or both.

In both light and heavy chains, the variable and constant domains are linked by a short segment of polypeptide chain, called the switch. Visual inspection of the Fab structures (for example, Fig. 2) identifies the switch peptides as residues 107, 108 and 109 in the light chains and residues 113, 114 and 115 in the heavy chains [numbering scheme of Kabat *et al.* (1991)]. Residue 106a is missing in κ chains so that the switch region in κ chains is shorter by one residue than those in λ chains and heavy chains. The relative disposition of the variable and constant modules is usually described by the angle between the V_L–V_H and C_L–C_H1 pseudo-dyads. This angle is called the "elbow bend" of the Fab and is variable, in view of the flexibility of the switch. The Fab bend ranges from a tight 127.2° (in Fab 8F5, PDB Entry: 1BBD) to an almost straight 176.2° (in Fab R19.9, PDB Entry: 2F19) (Table 3, Fig. 3). In all cases known to date where the Fab is bent, the Fd is more bent than the light chain, i.e. the angle between V_H and C_H1 is less than that between V_L and C_L; this has been attributed to the presence of smaller side chains in the interface between V_H and C_H1 (Segal *et al.*, 1974).

The Fv

On the basis of sequence variation, the residues in the variable domains are assigned either to hypervariable or complementarity-determining regions (CDRs), or to nonhypervariable or framework regions (Wu and Kabat, 1970). The CDRs of the light chain are defined as being comprised of residues 24-34 (CDR1-L), 50-56 (CDR2-L) and 88-97 (CDR3-L); those of the heavy chain contain residues 31-35 (CDR1-H), 50-65 (CDR2-H) and 95-101 (CDR3-H) [numbering convention of Kabat *et al.* (1991)]. Variations in length accompany the variability in sequence in these CDRs, with CDR3-H displaying particularly large length variations (Kabat *et al.*, 1991). The framework regions in V_L are defined as being comprised of residues 1-23 (FR1-L), 35-49 (FR2-L), 57-87 (FR3-L) and 98-107 (FR4-L); those in V_H contain residues 1-30 (FR1-H), 36-49 (FR2-H), 66-94 (FR3-H); and 102-112 (FR4-H).

In three-dimensions, the CDRs are seen as loops mainly situated at the N-terminal tip of the Fab (Fig. 4), where they form a continuous surface approximately 2800 Å^2 in area. The extent and conformation of each CDR are primarily determined by the nature and number of amino acids in the segment, and the variability in sequence and size seen in the CDRs results in a large variation in the topography of the CDR surface. The

Fig. 2. Stereodrawing of the α-carbon trace of the Fab of the murine antibody, HyHEL-10. The light chain is drawn with thinner lines. The variable domains are on top and the constant domains are at the bottom. The N- and C-termini of the two chains, residues in the switch regions, and the first and last residues of the six CDRs are labeled. The CDR residues are indicated by filled circles.

Best Available Copy

Table 3. Symmetry in the quaternary structure of murine and human antigen-binding regions of known three-dimensional structure

Antibody	PDB Code	V_L–V_H symmetry Rotation (degrees)	V_L–V_H symmetry Translation (Å)	C_L–C_H1 symmetry Rotation (degrees)	C_L–C_H1 symmetry Translation (Å)	Fab bend (degrees)
(Murine)						
J539	2FBJ	168.7	0.2	173.9	−2.0	143.8
McPC603	1MCP	173.4	0.1	171.8	−2.8	131.3
		173.6	0.3	171.9	−2.7	131.8[a]
HyHEL-10	3HFM	170.6	−0.3	167.7	−2.0	145.6
HyHEL-5	2HFL	171.4	0.1	169.1	−2.0	161.7
R19.9	2F19	175.2	0.3	166.8	−2.0	176.2
	1FAI	175.0	0.3	167.8	−1.8	176.1
4-4-20	4FAB	175.5	−0.3	174.3	1.6	174.9
BV04-01	[d]	176.8	−0.1	170.6	−1.8	173.5[a]
	[d]	173.5	0.3	172.3	1.3	172.5
36-71	6FAB	176.2	−0.1	171.5	−2.0	166.6
B13I2	2IGF	172.4	0.1	170.6	−2.0	155.8[a]
	1IGF	172.5	0.0	170.0	−2.0	153.0[b]
		172.7	0.1	170.9	−2.0	155.4[c]
D1.3	1FDL	166.2	0.7	171.1	−1.9	173.0
Yst9-1	1MAM	175.1	−0.1	172.4	−1.5	148.1
AN02	1BAF	173.3	−0.5	167.4	−2.2	153.8
17/9	1HIN	168.9	−0.3	172.3	1.9	175.1[a]
	1HIM	169.1	0.2	171.3	−1.7	172.7[a,b]
		169.5	−0.3	171.2	1.8	171.6[a,c]
	1HIL	170.3	−0.1	171.8	−1.9	159.7[b]
		168.9	0.2	171.7	2.0	160.5[c]
8F5	1BBD	173.7	−0.7	171.0	−2.1	127.2
NC41	1NCA	177.6	0.4	171.2	2.1	146.6
(Human)						
Kol	2FB4	167.9	−0.1	170.4	3.4	165.2
NEW	7FAB	164.3	−0.5	170.4	−3.8	129.9
Hil	8FAB	171.2	0.3	174.6	2.7	147.2[b]
		174.8	0.1	168.8	−3.3	135.7[c]
3D6	1DFB	166.8	0.2	169.5	−2.1	174.9
POT	1IGM	174.4	0.0	—	—	—

[a]Liganded form.

[b]For the first Fab in the asymmetric unit of the crystal.

[c]For the second Fab in the asymmetric unit of the crystal.

[d]Kindly provided by Dr Allen B. Edmundson and coworkers.

The rotation and translation parameters which relate V_L to V_H and C_L to C_H1 were obtained with program ALIGN (G. H. Cohen, NIH) using the α-carbon positions of residues: 3-7, 11-26, 31-38, 44-48, 60-66, 70-90 and 97-106 in V_L; residues: 3-7, 10-25, 32-39, 45-49, 65-71, 77-94 and 102-111 in V_H; residues 109-117, 129-142, 144-149, 158-162, 170-186, 192-198 and 204-208 in C_L; and residues: 119-127, 139-152, 154-159, 166-170, 177-193, 198-204 and 210-214 in C_H1. The numbering convention of Kabat *et al.* (1991) is used for V_L, V_H and C_L; Eu numbering (Edelman *et al.*, 1969) is used for C_H1. These residues represent the structurally equivalent positions in the homologous domains of J539 Fab (Suh *et al.*, 1986; Bhat, Padlan and Davies, in preparation). The Fab bend is the angle between the rotation axes which relate V_L to V_H and C_L to C_H1.

CDR surface is characterized by depressions and protrusions and may contain a deep pocket [for example, in McPC603 (Segal *et al.*, 1974), or cleft for example, in MAb131 (Garcia *et al.*, 1992)], or a protruberance [for example, in HyHEL-10 (Padlan *et al.*, 1989)]. Crystallographic analysis of several antibody–antigen complexes and other studies have repeatedly shown that antigen binding primarily involves this surface. The CDR surface is therefore usually equated with the combining site of the antibody (the paratope).

The six CDRs are disposed (Fig. 4) such that the N-terminal part of CDR1-L and the C-terminal parts of CDR2-L and CDR2-H are farther from the center of the CDR surface, while CDR1-H, CDR3-H, CDR3-L, the C-terminal part of CDR1-L, and the N-terminal parts of CDR2-L and CDR3-H are closer to the center.

Table 4. V_L–V_H and C_L–C_H1 interactions in murine and human antigen-binding regions of known three-dimensional structure

		V_L–V_H interactions					C_L–C_H1 interactions				
	PDB	Surface	Buried				Surface	Buried			
Antibody	Code	V_L	V_H	vdW	H.b.	I.p.	C_L	C_H1	vdW	H.b.	I.p.
(Murine)											
J539	2FBJ	773	740	190	9	0	858	835	147	3	0
McPC603	1MCP	874	825	170	6	0	765	750	120	1	0
		874	825	170	6	0	837	790	111	3	0
HyHEL-10	3HFM	718	676	122	3	0	945	981	153	6	1
HyHEL-5	2HFL	655	627	119	9	1	836	810	139	6	1
R19.9	2F19	827	846	168	9	0	872	835	159	5	0
	1FAI	744	751	153	5	0	868	861	165	6	0
4-4-20	4FAB	670	696	119	6	1	864	870	131	5	1
BV04-01		702	671	105	3	0	913	867[a]	149	7	1
		662	688	123	6	0	958	930	182	6	1
36-71	6FAB	726	729	131	4	0	1068	1056	217	17	2
B13I2	2IGF	715	682	132	8	0	1045	973[a]	145	4	1
	1IGF	756	731	163	11	1	876	836[b]	166	6	0
		786	752	159	7	0	868	911[c]	161	7	0
D1.3	1FDL	732	684	195	11	1	866	864	161	12	1
Yst9-1	1MAM	644	626	80	3	0	918	837	159	4	0
AN02	1BAF	642	615	120	7	0	1043	1000	188	11	2
17/9	1HIN	730	706	137	10	2	788	758[a]	162	5	0
	1HIM	760	742	132	6	1	799	761[a,b]	151	8	1
		740	730	143	9	1	834	796[a,c]	169	4	0
	1HIL	757	741	166	14	1	788	765[b]	163	6	1
		744	744	163	11	1	870	809[c]	172	8	1
8F5	1BBD	665	635	140	12	0	906	849	161	4	0
NC41	1NCA	614	631	141	5	0	1074	991	191	8	0
(Human)											
Kol	2FB4	800	790	162	4	0	894	816	125	9	2
NEW	7FAB	623	594	119	6	0	709	648	98	6	1
Hil	8FAB	583	592	132	9	1	676	623[b]	96	4	0
		714	694	153	6	0	829	755[c]	129	9	1
3D6	1DFB	716	683	133	2	1	841	888	116	6	0
POT	1IGM	687	687	130	9	1	—	—	—	—	—

[a]Liganded form.
[b]First Fab in the entry.
[c]Second Fab in the entry.

Two atoms are said to be in van der Waals' contact (vdW) if the distance between them is at most the sum of their van der Waals' radii plus 0.5 Å. Polar atoms are said to be hydrogen-bonded (H.b.) if the distance between them is at most 2.90 Å (taken to be the standard hydrogen-bond distance) plus 0.5 Å. Oppositely-charged atoms are said to form an ion pair (I.p.) if the distance between them is at most 2.85 Å (taken to be the standard ion-pair distance) plus 0.5. An average error in the atomic positions of 0.35 Å is assumed for all the structures included in the analysis; the assumed error in the interatomic distances is then 0.5 Å [= 0.35 × sqrt(2.0)]. Surface areas were computed using program MS of Connolly (1983); a probe radius of 1.7 Å was used and four points per square Ångstrom of surface area were computed.

The nonhypervariable or framework regions, by and large, show conserved amino acid substitutions and very similar three-dimensional structures. The different antibody combining sites, therefore, can be pictured as being constructed with CDRs of varied shapes and sizes, which are grafted onto a scaffolding of basically conserved structure.

The contact between V_L and V_H involves both framework and CDR residues and features framework–framework, framework–CDR and CDR–CDR interactions. The involvement of CDR residues in the V_L–V_H contact is significant, ranging from 26 to 57% of all atomic interactions in the structures analysed (Table 5). This involvement undoubtedly contributes to the variation seen in the quaternary association of the variable domains. The contribution to the contact from the individual CDRs is not the same (Table 6); in V_L, only CDR3-L interacts with all the CDRs of the heavy chain (on average, 5% of its contact is with CDR1-H, 28% is with CDR2-H and 67% is with CDR3-H); in V_H,

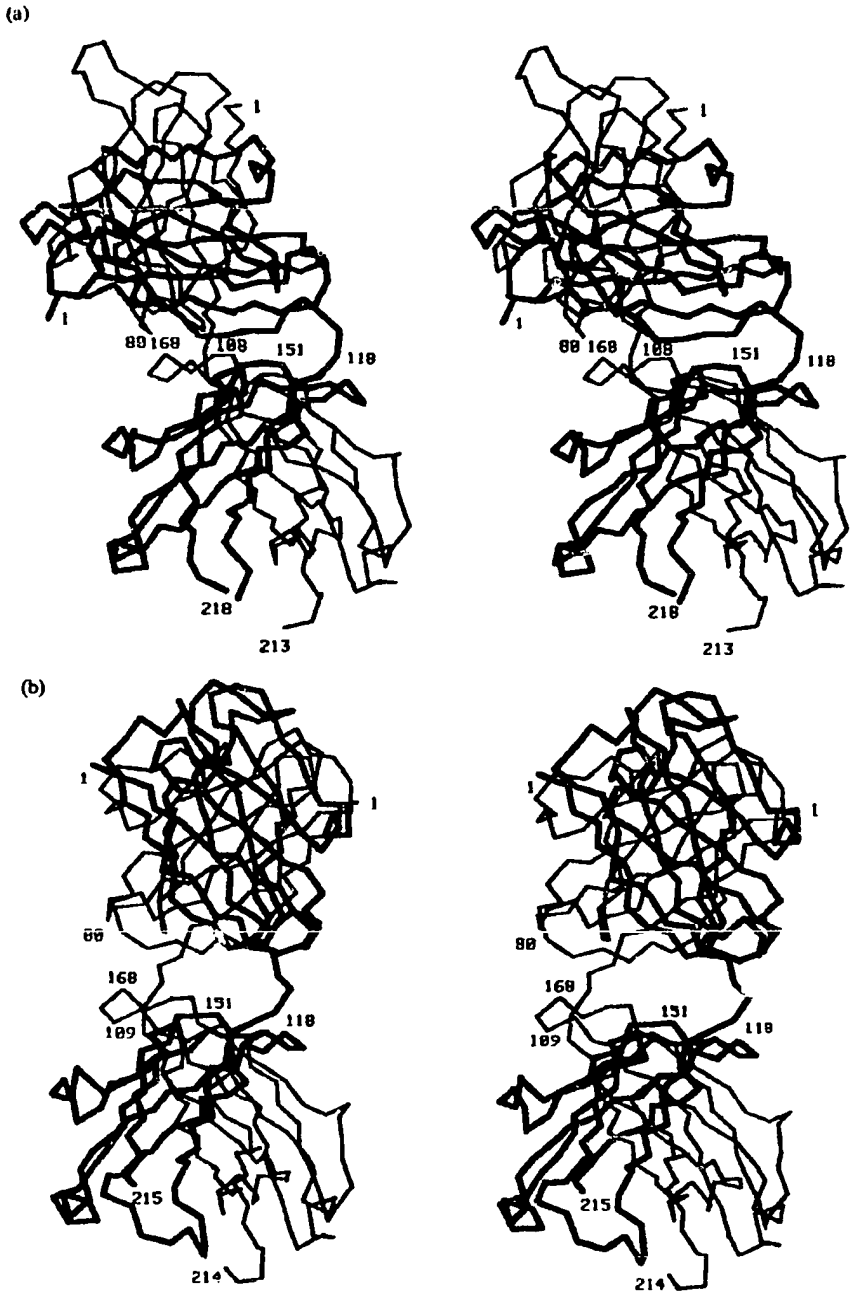


Fig. 3. Stereodrawings of the α-carbon trace of the Fabs of (a) 8F5 (PDB Entry: 1BBD), which is the most bent (Fab bend = 127.2°), and (b) R19.9 (PDB Entry: 2F19), which is almost straight (Fab bend = 176.2°). In these drawings, the Fabs are oriented such that their C_L–C_H1 modules are maximally superposed. The light chains are drawn with thinner lines and the heavy chains with thicker lines. The variable domains are on top and the constant domains are at the bottom.

only CDR3-H interacts with all the CDRs of the light chain (on average, 22% of its contact is with CDR1-L, 18% is with CDR2-L, and 60% is with CDR3-L). The framework residues of V_L which interact with the CDRs in the heavy chain are mostly from FR2-L and FR4-L, and occasionally from FR1-L; the framework residues of V_H which interact with the CDRs in the light chain are exclusively from FR2-H and, in almost all cases, the only residue that is involved is that at position 47 (usually a Trp). Details of the V_L–V_H contacts in Kol and J539 are presented in Table 7 and Fig. 5. The CDR and framework residues, which are involved in the V_L–V_H interaction in the various antigen-binding regions of known three-dimensional structure, are given in Table 8.

Intradomain framework–CDR interactions can influence the conformation of the CDRs. Indeed, canonical

Fig. 4. Stereodrawing of the α-carbon trace of the Fv of antibody HyHEL-10 viewed end on. The figure is rotated 90° relative to Fig. 2. The V_L is on the left (thinner lines). The CDR residues are indicated by filled circles and their ends are labeled.

structures have been observed for most CDRs and those structures appear to be determined by the nature of a small number of framework residues that interact with the CDRs (Chothia and Lesk, 1987; Chothia *et al.*, 1989; Tramontano *et al.*, 1990). The framework, therefore, does not simply provide a foundation for the

Table 5. Involvement of framework (FRM) and complementarity-determining (CDR) regions in the V_L–V_H interactions in murine and human antigen-binding regions of known three-dimensional structure

		Number of interatomic contacts between:				
Antibody	PDB Code	V_L and V_H FRMs	V_L FRM and V_H CDRs	V_H FRM and V_L CDRs	V_L CDRs and V_H CDRs	Total
(Murine)						
J539	2FBJ	65	36	24	65	190
McPC603	1MCP	56	27	27	60	170
		56	27	27	60	170
HyHEL-10	3HFM	51	8	17	46	122
HyHEL-5	2HFL	65	25	12	17	119
R19.9	2F19	50	27	17	74	168
	1FAI	48	40	15	50	153
4-4-20	4FAB	40	30	22	27	119
BV04-01	[a]	55	14	8	28	105
		49	32	15	27	123
36-71	6FAB	53	31	21	26	131
B13I2	2IGF	41	25	11	55	132
	1IGF[b]	68	26	11	58	163
	[c]	66	30	10	53	159
D1.3	1FDL	68	36	21	70	195
YST9-1	1MAM	47	11	13	9	80
AN02	1BAF	53	19	17	31	120
17/9	1HIN	54	26	18	39	137
	1HIM[b]	47	29	17	39	132
	[c]	50	24	17	52	143
	1HIL[b]	57	36	19	54	166
	[c]	57	33	15	58	163
8F5	1BBD	66	21	22	31	140
NC41	1NCA	58	17	15	51	141
(Human)						
Kol	2FB4	74	23	21	44	162
NEW	7FAB	51	14	22	32	119
Hil	8FAB[b]	59	23	12	38	132
	[c]	55	24	10	64	153
3D6	1DFB	49	14	16	54	133
POT	1IGM	55	28	11	36	130

[a]Liganded form.
[b]For the first Fab in the asymmetric unit of the crystal.
[c]For the second Fab in the asymmetric unit of the crystal.
See footnote to Table 4 for the definition of contacts.

Best Available Copy

Table 6. van der Waals' contacts among the CDRs of murine and human antigen-binding regions of known three-dimensional structure

	Interatomic contacts								
	CDR1-L with:			CDR2-L with:			CDR3-L with:		
Antibody	CDR1-H	CDR2-H	CDR3-H	CDR1-H	CDR2-H	CDR3-H	CDR1-H	CDR2-H	CDR3-H
J539	0	0	23	0	0	7	1	12	22
McPC603	0	0	15	0	0	2	13	4	26
HyHEL-10	0	0	5	0	0	0	0	33	8
HyHEL-5	0	0	0	0	0	0	4	9	0
R19.9	0	0	5	0	0	6	1	10	52
4-4-20	0	0	9	0	0	17	1	0	0
BV04-01	0	0	0	0	0	9	0	0	19
36-71	0	0	11	0	0	6	0	3	6
B13I2	0	0	36	0	0	3	0	0	16
D1.3	0	0	5	0	0	0	1	19	45
Yst9-1	0	0	0	0	0	4	0	3	2
AN02	0	0	2	0	0	0	0	26	3
17/9	0	0	9	0	0	10	0	4	16
8F5	0	0	0	0	0	16	1	14	0
NC41	0	0	3	0	0	6	0	0	42
Kol	0	0	5	0	0	0	3	8	28
NEW	0	0	4	—	—	—	3	12	13
Hil	0	0	10	0	0	1	0	1	26
3D6	0	0	2	0	0	18	1	0	33
POT	0	0	0	0	0	7	0	2	27
Totals	0	0	144	0	0	112	29	160	384

The coordinates used in the computations were for the liganded forms of McPC603, B13I2, BV04-01 and 17/9 (PDB Entry: 1HIM, first Fab in the entry); and (here and in subsequent calculations) those in PDB Entry 2F19 for R19.9, the first Fab in PDB Entry 8FAB for Hil, and those in PDB Entry 1NCA for NC41. The 7 residue deletion around the CDR2-L of NEW V_L is interpreted here as a deletion of all of CDR2-L in view of the minigene hypothesis of Kabat *et al.* (1978); this region in NEW has unusual amino acids and a different structure compared to the other V_Ls. See footnote to Table 4 for the definition of contacts.

construction of the combining site; small variations in the architecture of the framework can affect the topography of the CDR surface.

Crystallographic analysis of antibody structures, with and without bound ligand, has shown that, in some cases, the combining site structure could change on

Table 7a. V_L–V_H contacts in the human (IgG1,λ) Kol (PDB Entry: 2FB4)

	V_L																	
V_H	T32	N34	Y36	Q38	G41	M42	A43	P44	L46	Y49	Y87	W91	V93	N95a	A95b	Y96	F98	Totals
Y35																3		3
V37																	4	4
Q39				6[b]	2						1							9
G44											4							4
L45				2				2			7						9	20
W47														1	6	14	5	26
I50												3						3
H58														5				5
F91					2	3	2											7
C100a												7						7
S100b												2						2
S100c													2					2
C100f	1											6				2		9
F100g		3[b]								10								13
G100h		1														3		4
P100i			10[b]													6		16
D101									3									3
W103			5				1	15									2	23
G104							2											2
Totals	1	4	15	8	4	3	5	17	3	10	12	18	2	6	6	28	20	162

Table 7b. V_L–V_H contacts in the murine (IgA,κ) J539 (PDB Entry: 2FBJ)

V_H	V_L S32	H34	Y36	Q38	S43	P44	P46	Y49	E50	A55	Y87	Q89	W91	Y93	P94	L95	I96	F98	Totals
W33														1					1
V37																		4	4
Q39				9[h]							1[h]								10
K43											1								1
G44											3								3
L45						3					5							6	14
W47															2	1	21	1	24
E50														1					1
T56														1					1
N58														5[h]	4				9
P61																1			1
Y91				1	4	3													8
L95																	1		1
Y97									1										1
Y98	4	6[h]							5				11						26
G99		3										3[h]	4				2		12
Y100		10	5				3	16		1									35
N100a			6[h]				2					1						1	10
A101							3												3
W103			3			6	3											1	13
G104					3[h]														3
Q105					8[h]														8
Totals	4	19	14	10	15	12	11	16	6	1	10	4	15	8	6	2	24	13	190

The residues on the horizontal lines are from the V_L and those along the vertical are from the V_H. In these matrices of contacts, the element $c(i,j)$ represents the number of interacting atom pairs, one from residue i and the other from residue j. The superscript [h] that is found after some matrix elements signifies that the contact involves at least one hydrogen bond. See footnote to Table 4 for the definition of contacts.

Fig. 5. Stereodrawings of the α-carbon trace of the Fvs of Kol (a) and of J539 (b) showing the residues which are in the V_L–V_H contact. The V_Ls are drawn with thinner lines on the left. The side chains of the residues involved in the contact are shown in full.

Best Available Copy

Table 8a. V_L residues involved in the V_L–V_H contact in murine and human antigen-binding regions of known three-dimensional structure

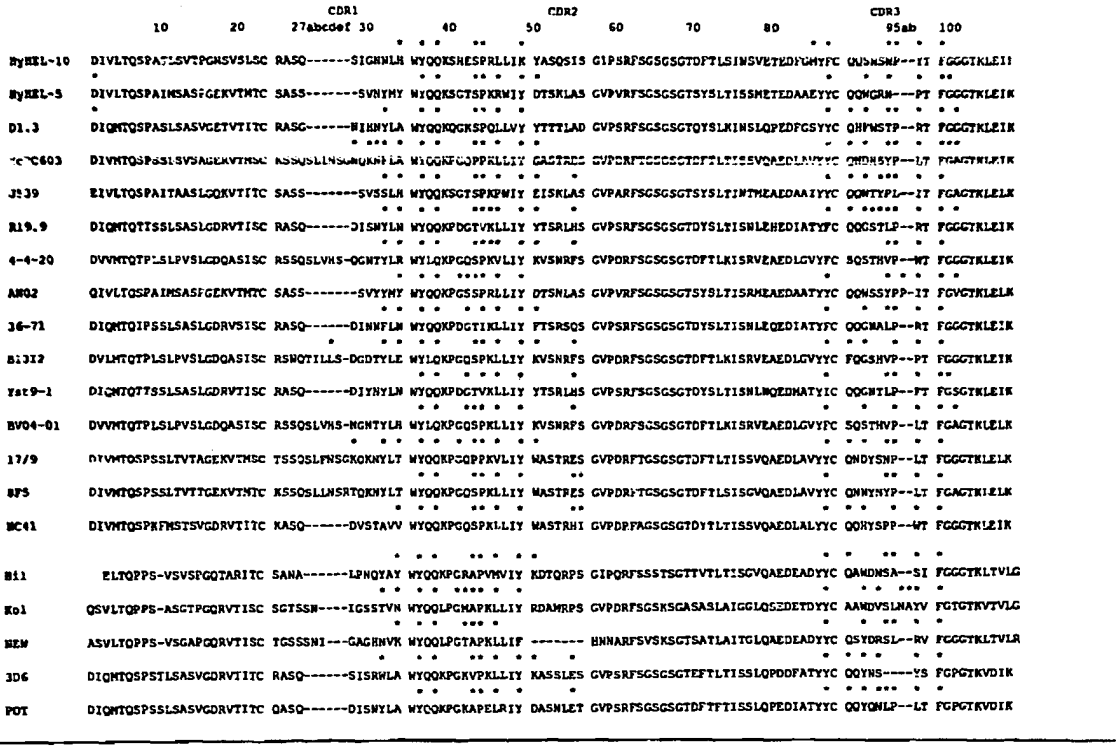
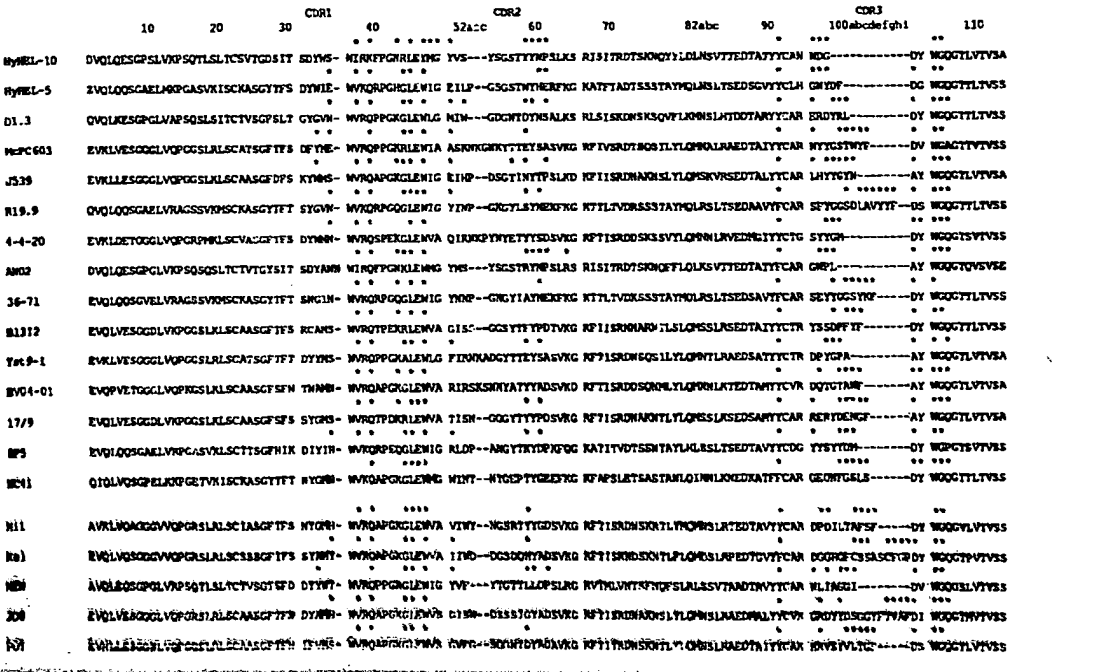


Table 8b. V_H residues involved in the V_L–V_H contact in murine and human antigen-binding regions of known three-dimensional structure



The residues which contact the opposite domain are indicated by (*). See footnote to Table 4 for the definition of contacts.

binding, in the manner of an "induced fit" {Edmundson *et al.*, 1974, 1987; Colman, 1988; Bhat *et al.*, 1990; Herron *et al.*, 1991; Rini *et al.*, 1992; [reviewed by Davies and Padlan (1992)]}. The conformation of the individual CDRs, especially the longer ones, could change, as could the mode of quarternary association of the variable domains (Colman *et al.*, 1987; Colman, 1988; Bhat *et al.*, 1990; Herron *et al.*, 1991; Rini *et al.*, 1992). These results suggest that the combining site structure is not rigid, rather, it is plastic and may assume different conformations depending on circumstance.

The C_L–C_H1 module

Representative C_L–C_H1 interactions are presented in Table 9 and Fig. 6. The contacts observed in Fab Kol (human IgG1,λ) [Table 9a and Fig. 6(a)] include two salt bridges: one between Glu123-L and Lys213-H and the other between Glu124-L and Lys147-H. In Fab 3D6 (human IgG1,κ) (Table 9b), no salt bridges are formed, despite the fact that a glutamic acid residue is present at position 123 also in the 3D6 light chain. Actually, in 3D6, there is a favorable electrostatic interaction between the charged moieties of Glu123-L and Lys213-H, which are within 4.0 Å of each other. This interaction would be only about half as strong as that in Kol where the corresponding distance is 2.8 Å. The second salt bridge in Kol is not possible in 3D6 because the Glu at position 124 in C_λ is replaced by a Gln in C_κ (Table 2). Despite the many differences in sequence, the patterns of C_L–C_H1 interaction in these human Fabs, one with a λ and the other with a κ chain, are similar.

The Fab of 17/9 (IgG2a,κ) (PDB Entry: 1HIL, first Fab in the entry) was chosen to represent the murine IgG subclasses in terms of the C_L–C_H1 interactions, since this structure has been determined to high resolution and refined to a high degree (Table 1). The C_L–C_H1 contact in 17/9 (Table 9c) shows a pattern similar to that of 3D6 (Table 9b), which is not surprising in view of the similarity in sequence of the homologous domains. The salt bridge between Glu123-L and Lys213-H is also found in 17/9 and in the C_L–C_H1 of the other murine IgG isotypes (data not shown). The inter-residue contacts in the C_L–C_H1 of the murine J539 (IgA,κ) Fab (PDB Entry: 2FBJ) are presented in Table 9d and Fig. 6(b) Again, the pattern of the contacts is similar to the others. However, the salt bridge found in the C_L–C_H1 interface in the murine and human IgGs is not possible in IgAs (murine or human), since the Lys at heavy-chain position 213 is not present in these molecules.

A cavity has been observed between C_L and C_H1 (Padlan *et al.*, 1986). It was proposed that the function of this cavity is to provide greater flexibility in the interaction between the domains, so that the variation in the interface residues in the different isotypes can be tolerated while still preserving the basic quaternary structure of the C_L–C_H1 module.

The Fab bend

The interactions between V_L and C_L and between V_H and C_H1 are not very strong and, in view of the variation in the Fab bend (Table 3), are variable. The contacts, computed for the most bent Fab, 8F5 [Fig. 3(a)], and for the most straight Fab, R19.9 (PDB Entry: 2F19) [Fig. 3(b)], are presented in Table 10. Included in Table 10 are the contacts for an Fab with an intermediate bend angle (153.0°), B13I2 (PDB Entry: 1IGF, first Fab in the entry).

In the case where the Fab is most bent [Fig. 3(a), Table 10a], the residues from V_H, which are involved in the contact, are from the N- and C-terminal segments of the domain; those from C_H1 are from the N-terminal segment and from two bends: one at around position 151 and the other at around position 205; those from V_L come from the bend at around position 81 and from the C-terminal segment of the domain; those from C_L come from two bends: one at around position 140 and the other at around position 168. As the Fab becomes straighter (Table 10b), fewer contacts are formed; the V_H and C_L segments involved in the interaction remain in contact, but interactions involving the N- and C-terminal segments of C_H1 become less and less and are ultimately lost, as well as the interactions involving the bend at around position 81 of V_L. Not surprisingly, as the Fab becomes almost straight [Fig. 3(b), Table 10c], the N-terminal segment of V_L starts to be involved in the contact, so that the V_L–C_L and V_H–C_H1 interactions become more analogous.

The biological function of the C_L–C_H1 module is not obvious although the association of C_L and C_H1 necessarily increases the probability of a proper V_L–V_H quaternary interaction. In addition, the presence of additional domains in the Fab arms extends the reach of the antibody and the loose connection between the Fv and the C_L–C_H1 module contributes to the flexibility of the molecule (see below). It is also conceivable that the C_L and C_H1 may help conceal other ligand binding sites that are uncovered only when antigen is bound. A more definite function for C_L–C_H1 may be revealed in time.

The Fc

The structure of this fragment has been reviewed elsewhere (Padlan, 1990*b*) and some parts of that review are reproduced here. Crystal structures are available for the Fc of human IgG1 (Deisenhofer, 1981) and for that of rabbit IgG (Sutton and Phillips, 1983), and they are very similar. However, atomic coordinates are available for only the human Fc (PDB Entries: 1FC1 and 1FC2). The structure of the pFc′ of guinea pig IgG has also been analysed (Bryant *et al.*, 1985) (PDB Entry: 1PFC) and it is found to be similar to that of the Cγ3–Cγ3 module of human Fcγ. We will confine our discussion to the structure of human Fcγ.

The structure that is available for human Fcγ is comprised of residues 238-443 (Eu numbering), which includes most of the Cγ2 and Cγ3 domains. In some crystal forms of Fc or of intact antibody, the two halves of the fragment are related by a crystallographic, i.e. exact, two-fold axis. In the crystal form of human Fcγ that was analysed crystallographically, the two chains are related by a pseudo-dyad axis. In that structure

Table 9a. C_L–C_H1 contacts in the human (IgG1,λ) Kol (PDB Entry: 2FB4)

	V125	F126	P127	L128	A129	K133	A141	L142	L145	K147	H168	F170	P171	V173	Q175	S176	L182	S183	V185	K213	K218	S219	C220	Totals
T116							1																	1
F118				9	2		5	2											2					20
S121		2	3																					5
S122																					1			1
E123	1	4	1																	6[a]				12
E124		13								4[a]														17
K129										2														2
T131									2	2														4
V133									2									1						3
L135												7						2	1					10
I136												2												2
S137											1	2												3
E160														2	4	3								9
T162														3										3
S165													1											1
Q167											3													3
A174												2												2
A175												2												2
S176												2												2
Y178									1					3			2	5						11
T208						2																		2
E213																					4		2	6
C214																						4		4
Totals	1	19	4	9	2	2	6	2	5	8	4	17	1	8	4	3	2	8	3	6	5	4	2	125

Table 9b. C_L–C_H1 contacts in the human (IgG1,κ) 3D6 (PDB Entry: 1DFB)

	F126	P127	L128	A129	S131	S132	S136	T139	A141	L145	K147	H168	F170	P171	V173	Q175	V185	T187	S219	Totals
F116						6	5	2	6									2		21
I117						6														6
F118			7	1	1				5											14
S121		3																		3
E123		2																		2
Q124	13										1									14
T129											1									1
S131										1	3									4
V133			3																	3
L135									1				1				3			5
N137												2						1		3
N138												1								1
Q160															1	3				4
S162													1	8	1					10
V163														3						3
T164												1	4							5
S174												1	4							5
L175													3							3
S176													7							7
T180											1									1
C214																			1	1
Totals	13	5	10	1	1	12	5	2	12	1	6	5	20	11	2	3	3	3	1	116

Table 9c. C_L–C_H1 contacts in the murine (IgG2a,κ) 17/9 (PDB Entry: 1HIL, first Fab in the entry)

	Y126	P127	L128	A129	P130	V131	T135	G137	L145	K147	H168	T169	F170	P171	V173	Q175	T181	S183	S184	S185	K213	Totals
S116							4															4
F118			12	4	2		4															22
P119						1																1
S121	3	4																				7
E123	2	2																			2[a]	6
Q124	19																					19
S131									2	3												5
V133			4						1													5
F135			2				1	2					4					2	2	8		21
N137													1							3		4
N138											5											5
L160															4	2	2					8
N161															1							1
S162													4	8	1							13
W163														4								4
T164												1	2									3
D167											2											2
S174											6		4									10
M175													8									8
S176													9					3				12
S180										3												3
Totals	24	6	18	4	2	1	9	2	3	6	13	1	32	12	6	2	2	5	2	11	2	163

Table 9d. C_L–C_H1 contacts in the murine (IgA,κ) J539 (PDB Entry: 2FB[illegible])

	Y126	P127	L128	T129	L130	L134	I141	L145	H147	T166	V168	F170	P171	A173	L174	A175	T181	S183	Q185	Totals
S116							1													1
I117					1	2														3
F118			7	6	6		6													25
S121		6																		6
E123	4	3																		7
Q124	18																			18
S127	2																			2
S131								2	1											3
V133			2					1												3
F135			2									6						3	10	21
N137																			8	8
L160														4	4	4	2			14
N161														1						1
S162													2	1						3
W163													3							3
T164											1	2								3
K169										1										1
S174											1	3								4
M175												7								7
S176												10						2		12
T178								1												1
T180									1											1
Totals	24	9	11	6	7	2	7	4	2	1	2	28	5	6	4	4	2	5	18	147

The residues on the horizontal lines are from C_L and those along the vertical are from C_H1. In these matrices of contacts, the element $c(i,j)$ represents the number of interacting atom pairs, one from residue i and the other from residue j. The superscript [a] that is found after some matrix elements indicates that the contact includes a favorable electrostatic interaction; the superscript [b] signifies that the contact involves at least one hydrogen bond. In 3D6 and in Kol, a disulfide bond exists between Cys214-L and Cys220-H. See footnote to Table 4 for the definition of contacts.

Best Available Copy

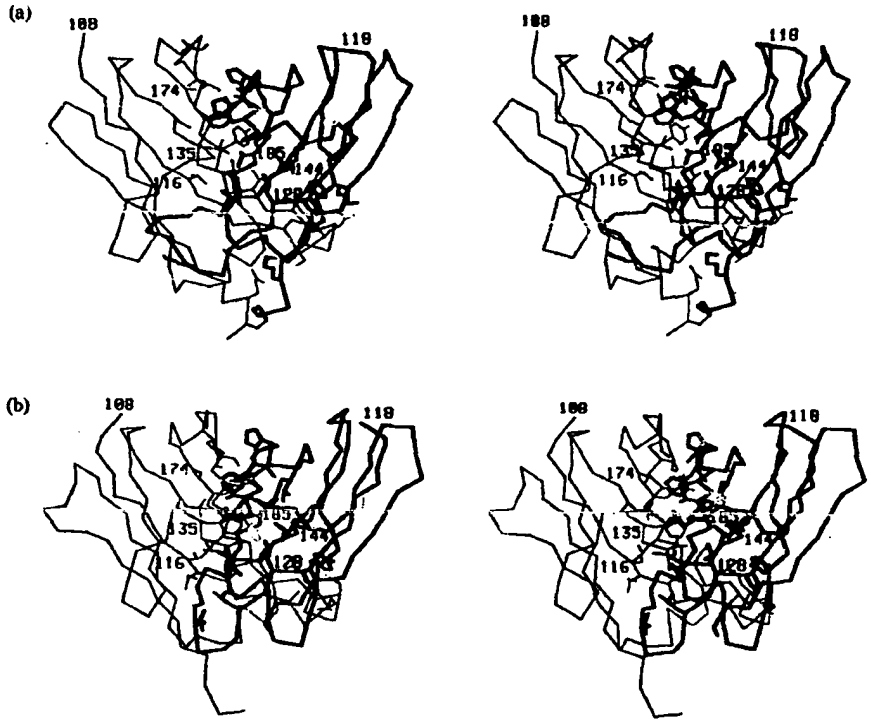


Fig. 6. Stereodrawings of the α-carbon trace of the C_L–C_H1 modules of Kol (a) and of J539 (b) showing the residues which are in the C_L–C_H1 contact. The C_Ls are drawn with thinner lines on the left. The side chains of the residues involved in the contact are shown in full.

(Fig. 7), the two Cγ3 domains are related by a rotation angle of 179.3° while the Cγ2 domains are related by a rotation of 174.3°. Carbohydrate moieties, that are linked to the asparagines at position 297 (Eu numbering) in both chains, lie between the two Cγ2 domains. The Cγ3 domains are in close association and form a compact globule, while the Cγ2 domains are farther apart.

The Cγ2 and Cγ3 domains both resemble closely the constant domains of the Fabs. The residues which form the bilayer β-pleated sheet structure are indicated in Table 2. The interior of both domains is filled with mainly hydrophobic side chains. Side chains emanating from one β-sheet in each bilayer form the contact between the Cγ3 (Fig. 8). The interaction is strong and involves more than 20 residues from each domain (Table 11). Approximately 2000 Å^2 of surface area are buried in the Cγ3–Cγ3 interface, where several hydrophobic residues, including some with large, aromatic side chains, are found. In addition, there are a number of hydrogen-bond interactions and three salt bridges.

In contrast, the analogous sheets in the Cγ2 domains are covered by the carbohydrates linked to the asparagines at 297 (Fig. 7), so that a mode of association, like the one in the Cγ3 domains, is not possible. Instead, the Cγ2 domains interact through their carbohydrate moieties, and only weakly. Even without the

Table 10a. Atomic contacts between the variable and constant domains of the same chain in 8F5 Fab (PDB Entry: 1BBD)

			C_H1					
V_H	A118	K119	T120	F150	P151	P153	P205	A206
A9						1	2	1
L11	1	1	4	1	1			
S108						1		
T110				1	3			
S112				2				

			C_L		
V_L	Y140	Q166	S168	K169	S171
A80			5	2	
E81			2		
L83		4			
E105		2			
I106	2	11			1
K107	1				

Table 10b. Atomic contacts between the variable and constant domains of the same chain in B1312 Fab (PDB Entry: 1IGF)

		C_H1		
V_H	A118	T120	F150	P151
L11	1	1	3	3
T110			1	
S112			2	

			C_L		
V_L	Y140	Q166	S168	S171	Y173
L83		2	1		
E105	6	5			5
I106	1	6		4	
K107	5				

Table 10c. Atomic contacts between the variable and constant domains of the same chain in R19.9 Fab (PDB Entry: 2F19)

	C_H1	
V_H	F150	P151
L11	4	1
S112	4	

	C_L			
V_L	Y140	Q166	S171	Y173
S12	1			
E105	4	3		4
I106	4	11	2	
K107	10			

The residues on the horizontal lines are from the constant domains and those along the vertical are from the variable domains of the Fabs. See footnote to Table 9.

carbohydrates, the $C\gamma2$ domains may not be able to associate in the manner of the $C\gamma3$; the N-termini of the $C\gamma2$ domains are probably too close, in view of the interchain disulfide bridge(s) in the hinge region, while the N-termini of the $C\gamma3$ domains are far (more than 40 Å) apart.

The longitudinal contact between the $C\gamma2$ and $C\gamma3$ domains is substantial, with approximately 780 Å^2 of surface area buried by the interaction. Seventeen residues, 8 from $C\gamma2$ and 9 from $C\gamma3$, are involved in this contact, which includes two salt-bridges (Table 12).

The amino acid sequence of human IgG1 $C\gamma2$ and $C\gamma3$ and those of the homologous domains of the other human heavy chains can be compared in Table 2. The various IgG isotypes are similar to a very high degree and there is little doubt that the structures of their Fc will be essentially the same. The sequences of the other heavy chains are not as similar but they share enough structural features with IgG1 to justify the assertion that the homologous structures in their Fc will have the same general architecture.

No obvious structural function can be attributed to the carbohydrate attached to the Asn at 297 in IgG1 (Fig. 7). Interestingly, all the other heavy chains are probably glycosylated at the same (homologous) position, except the IgAs. In the latter, a probable site of carbohydrate attachment is the Asn at position 258 (Table 2). It is interesting that the residue at position 258 in IgG1 is close to the end of the carbohydrate moiety (Fig. 7) and contacts the last sugar residue. One wonders whether, in the IgAs, a carbohydrate attached to residue 258 might follow a course reversed in relation to that found in IgG1, but nevertheless covering one face of C_H2 in much the same way as that observed in IgG1 (Deisenhofer, 1981). It should be pointed out that the

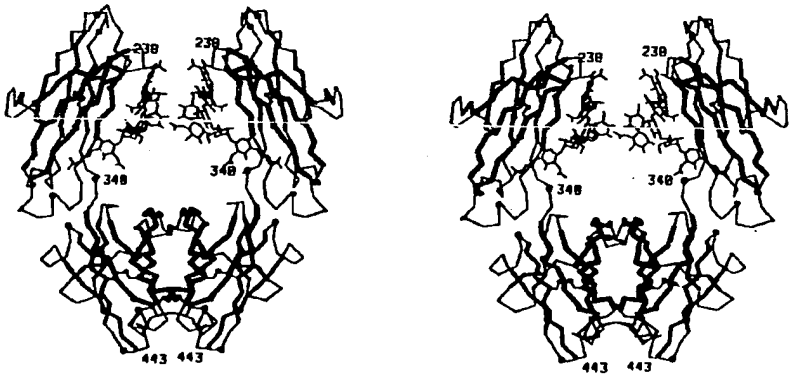


Fig. 7. Stereodrawing of the α-carbon trace of human IgG1 Fc. The C_H2 domains are on top and the C_H3 domains are at the bottom. The first (number 238) and last (number 443) residues that were visible in the crystal structure (Deisenhofer, 1981), and the C-terminus of the C_H2 domain (number 340) are labeled in each chain. Every tenth residue, starting from number 240, is indicated by a larger, open circle. The intradomain disulfide bonds are drawn with thicker bonds and filled circles in the middle of each domain. The β-pleated sheets are drawn with thick bonds, thicker for the first sheets.

Fig. 8. Stereodrawing of the α-carbon trace of C_H3–C_H3 module of human IgG1 Fc. The side chains of the residues involved in the interdomain contact are shown in full.

Table 11. C_H3–C_H3 contacts between the first and second chains in human IgG1 Fc (PDB Entry: 1FC1)

	2nd																						
1st	Q347	Y349	L351	P352	S354	E356	E357	T366	L368	K370	N390	K392	T394	P395	V397	L398	D399	S400	F405	Y407	K409	K439	Totals
Y349					1	5	12																18
L351			2	1	1																		4
P352			1																				1
S354		1	3																				4
E356		4																				5[a]	9
E357		10								3[a]													13
K360	2	1																					3
S364									1	1													2
T366			1																	7			10
L368																					2		2
K370							2																2
N390																		1					1
K392																4	1		5				10
T394													5		3				1				9
P395														1	1								2
V397													3	1									4
L398												3											3
D399												5									3		8
S400											1												1
F405												5	2								2		9
Y407								7					1							16	7		31
K409									1								2		3	6			12
K439					1	10[a]																	11
Totals	2	16	7	1	3	15	14	7	2	4	1	13	11	2	4	4	3	1	9	31	14	5	169

The residues on the horizontal line are from the second chain and those along the vertical are from the first chain of the human IgG1 Fc (PDB Entry: 1FC1). See footnote to Table 9.

Table 12. C_H2–C_H3 contacts in human IgG1 Fc (PDB Entry: 1FC1, first chain in the entry)

	C_H3									
C_H2	Y373	P374	E376	I377	E380	M428	H429	E430	H435	Totals
P247			1	1						2
K248					3*	5				8
L251						2	4	3	7	16
M252						3				3
L314								4	1	5
K338		1						7*		8
A339	2	1								3
K340	2									2
Totals	4	2	1	1	3	10	4	14	8	47

The residues on the horizontal line are from the C_H3 domain and those along the vertical are from the C_H2 domain of the first chain of the human IgG1 Fc (PDB Entry: 1FC1). See footnote to Table 9.

carbohydrate moieties attached to the asparagines at 297 in rabbit IgG were found to be asymmetrically disposed and to be in greater contact than those in human Fcγ (Sutton and Phillips, 1983). A detailed comparison of these two Fcγ structures, when atomic coordinates for the rabbit Fc become available, may reveal other differences.

The hinge

Antibodies exhibit segmental flexibility, which is made possible by the presence of the hinge region between the Fabs and the Fc and of the switch regions within each fragment [see, for example, Burton (1985, 1990*b*) and Tan *et al.* (1990)]. The hinge permits the fragments to rotate or to wag, while the switch regions permit the fragments to flex. By and large, the hinge can be viewed as consisting of three parts: a flexible upper region which permits the Fabs to rotate and wag and which also determines the separation between the Fab arms, a stiff middle which acts as a spacer between the Fabs and the Fc, and a flexible lower region which allows the Fc to wag. The rigidity of the middle part is almost certainly provided by the inter-heavy chain cystines and by the many proline residues that are often found in the hinge. The lower part of the hinge would be the segment between the last inter-heavy chain cystine and the residue which marks the beginning of the compact Fc (Pro238 in human IgG1 or homologous residue in the other antibody classes). The residues which constitute the flexible upper part of the hinge are not easily identified in the absence of three-dimensional structure. The size of the three hinge parts varies among the different antibody classes, so that the middle part, for example, ranges in size from the single cystine in human IgD to more than 40 residues in IgG3 (Table 2).

Very little three-dimensional information is available for the hinge, mainly because of segmental flexibility. The crystal structure of the intact human IgG1, Kol, for example, did not reveal the Fc, which ostensibly assumed several different positions and orientations and, thus, was "averaged out" in the electron-density map (Marquart *et al.*, 1980). The portion of Kol that was visible extended only to Cys229 (Eu numbering). Consequently, the octapeptide, PAPELLGG, between Cys229 (the last residue visible in Kol) and Pro238 [the first residue visible in the Fc structure (Deisenhofer, 1981)], has not been visualized and probably represents the most flexible part of human IgG1. The crystal structure of a murine IgG2a monoclonal antibody showing a complete hinge has been reported (Harris *et al.*, 1992), but full details of the structure have not been presented.

In the picture of the human IgG1 hinge that is available (Fig. 9), the segment, DKTHTCPPC, which represents the upper and middle portions of the hinge, contains strong secondary structural elements; the KTHT region is α-helical and the CPPC segments from the two chains form a poly-L-proline double helix, cross-linked by the two pairs of cysteines (Marquart *et al.*, 1980). This part of the IgG1 can be expected to be rigid. Indeed, a visual examination of this region in the crystal structure (Fig. 9) strongly suggests that at least a partial unraveling of the α-helical structure may

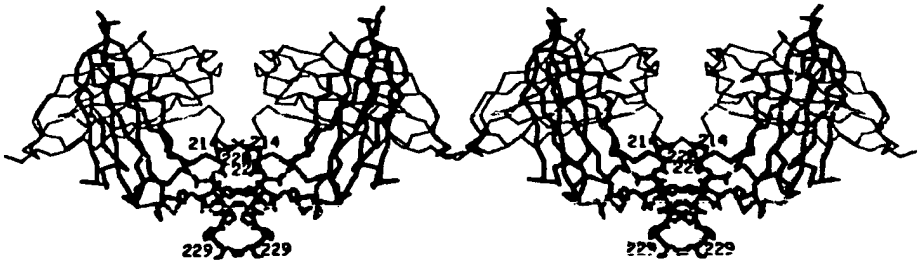


Fig. 9. Stereodrawing of the part of the hinge that is visible in the crystal structure of Kol (Marquart *et al.*, 1980). The orientation is the same as in Fig. 1.

be required to give the Fabs the freedom to rotate. In this antibody, the Fabs are separated by only 10 Å (the distance between the C-termini of the Fabs).

We can only speculate on the nature of the hinge in the other antibody classes and isotypes. The hinge of human IgD is probably the most floppy, having 46 amino acids between the putative end of the Fab and the inter-heavy chain cystine (Table 2); of course, the possible occurrence of helical structures in this stretch cannot be ruled out. The human IgD hinge also may provide the greatest possible separation between the Fabs [each of the two 46-a.a. segments could be as long as 160 Å (3.5 Å per residue) if fully extended]. The hinge of human IgG3, on the other hand, has been proposed to separate the Fabs from the Fc with a stiff rod 140 Å long in the form of a polyproline double helix stabilized by many disulfide bonds (Marquart *et al.*, 1980). The hinge regions of human IgG4 and of the two IgAs are probably also stiff in view of the presence of many prolines (Table 2).

The C_H2 domains of IgE and IgM could also be viewed as rigid Fab separators and thus function in part as "hinges". Modeling of the Fc of IgE (Padlan and Davies, 1986; Pumphrey, 1986; Helm *et al.*, 1991) positions the ends of the Fabs approximately 33 Å apart. A flexible segment following the (last) inter-heavy chain disulfide bridge has been proposed for IgE (Helm *et al.*, 1991) and for IgM (Perkins *et al.*, 1991).

Nature seems to have made provision for coping with different varieties of antigens by the use of the hinge. The variation in hinge length and amino acid sequence in the different heavy chain classes and isotypes results in antibodies with different reach and rotational adaptability, giving the immune system the means to cope with the possibility of different spacings and orientations of antigenic determinants (e.g. Beale and Feinstein, 1976; Burton, 1990*b*). One wonders also whether the variation in the structure of the hinge may have a bearing on the respective roles that the different antibody types play in the overall immune response.

The demonstrated flexibility of the lower portion of the hinge poses an intriguing question since a concerted conformational change in the two disulfide-linked chains would be required to produce an overall deformation in this part of the antibody. In view of the fact that alterations in the conformation of a polypeptide backbone are accomplished by changes in the peptide dihedral angles, which are restricted (Sasisekharan, 1962), this part of the hinge may not be dynamically, i.e. continuously, deformable. Indeed, flexed forms of antibodies, e.g. states in which the Fc and the Fabs are not co-planar, may persist for sufficiently long times to be observable (Zheng *et al.*, 1991, 1992).

STRUCTURAL ASPECTS OF ANTIBODY FUNCTION

Antigen binding

The antibody function for which we have the most information is antigen binding. It is well known that the binding of antibody to its antigen is exquisitely specific. Yet, it is also known that many antibodies are polyreactive, i.e. they are capable of binding to several different antigens, although with low affinity [see, for example, Burastero *et al.* (1988)]. Polyreactivity is usually associated with natural or pre-immune antibodies which frequently display reactivity towards self-antigens and which are usually encoded by unmutated or essentially unmutated germline genes [see, for example, Baccala *et al.* (1989) and Logtenberg (1990)]. The structural basis for polyreactivity is not yet clear. What is better understood is the basis for high-affinity antibody–antigen interactions.

(i) *Structural correlates of antigen-binding specificity.* The high specificity of the binding of antibody to its antigen is due to the complementarity in the structures of the antibody combining site and the antigenic determinant (the epitope). This is illustrated by the complexes between the antibodies D1.3, HyHEL-5 and HyHEL-10 and their antigen, hen egg white lysozyme (Table 13,

Table 13a. Contacts between antibody and ligand in the D1.3–lysozyme complex (PDB Entry: 1FDL)

	Antibody																
	Light Chain							Heavy Chain									
Lysozyme	[Y32	Y49	Y50	T53	F91	W92	S93]	[G31	Y32	W52	G53	D54	D97	Y98	R99]	Totals
D18			5													5
N19		1	4	2												7
G22		1											1		3	5
Y23													2			2
S24													9	2		11
N27													3			3
K116								2	1							3
G117								3		1	6	1				11
T118										1		4				5
D119										12				4		16
V120														4		4
Q121	9				4	8	1							7		29
I124	1					2										3
R125						6										6
Totals	10	2	9	2	4	16	1	5	1	14	6	5	15	17	3	110

Table 13b. Contacts between antibody and ligand in the HyHEL-5–Lysozyme complex (PDB Entry: 2HFL)

	Antibody																				
	Light Chain								Heavy Chain												
Lysozyme	[N31	Y32	Y34	D50	W91	G92	R93	P95]	[W33	E35	W47	E50	L52	S54	G55	S56	T57	N58	G95	N96	Y97]	Totals
Q41														2	1	3						6
T43									2							6	2	7				17
N44																		2				2
R45					11	4	12	5			3	4[a]										39
N46						2	3															5
T47						1	2															3
D48	2																					2
G49					3																	3
T51									1													1
Y53									5													5
G67																					5	5
R68					7				9	1[b]		2[b]							4	1	6	30
T69																					5	5
P70		5	1	3																	7	16
S81													1									1
L84													1	3		1						5
Totals	2	5	1	3	21	7	17	5	17	1	3	6	2	5	1	10	2	9	4	1	23	145

Table 13c. Contacts between antibody and ligand in the HyHEL-10–lysozyme complex (PDB Entry: 3HFM)

	Antibody																				
	Light Chain										Heavy Chain										
Lysozyme	[G30	N31	N32	Y50	Q53	S91	N92	W94	Y96]	[T30	S31	D32	Y33	Y50	S52	Y53	S54	S56	Y58	W95]	Totals
R14		1																			1
H15		8																			8
G16	3	6	1																		10
Y20			3			5	4														12
R21							4	2	3					6					6		21
W63													1			5					6
R73										12	4										16
L75											2	1				1					4
T89					3																3
N93				4	7																11
K96		4	1	13																	18
K97												1	8							1	10
I98													3								3
S100													1	5					1	4	11
D101													3		7	19	2	5	3		39
G102																		2	5		7
Totals	3	19	5	17	10	5	8	2	3	12	6	2	16	11	7	25	2	7	15	5	180

Table 13d. Contacts between antibody and ligand in the B13I2–peptide complex (PDB Entry: 2IGF)

	Antibody																		
	Light Chain						Heavy Chain												
Peptide	[D28	D30	Y32	G91	V94	P95]	[R31	A33	I51	S52	S52a	G53	S55	Y56	F58	Y95	P99	F100]	Totals
E1										2	3	5	7	3					20
V2							4			3	12						2		21
V3										1				3	2				6
P4								1	4	3					2	3	1		16
H5				4	6	5										8	2	6	31
K6					1										4				5
K7	6[a]	2[a]	4															2	14
Totals	6	2	4	4	7	5	4	1	4	9	15	5	7	6	8	13	5	8	113

The residues on the horizontal line are from the antibody and those along the vertical are from the ligand. See footnote to Table 9.

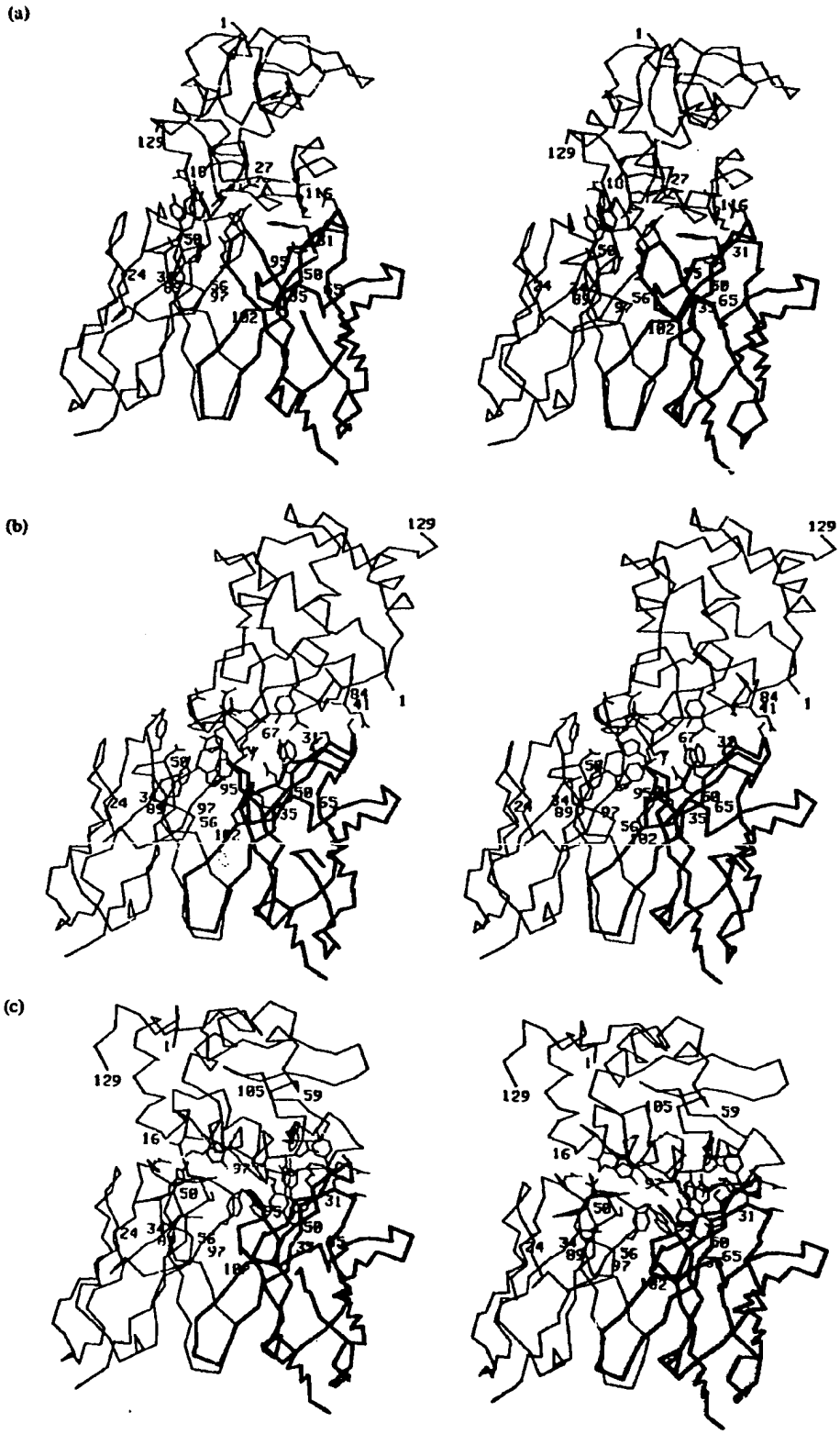


Fig. 10. Stereodrawing of the interaction between the antigen, hen egg white lysozyme, and the antibodies D1.3 (a), HyHEL-5 (b) and HyHEL-10 (c). Only the α-carbons are traced, except for the residues involved in the interaction, which are drawn in full. In these drawings, the lysozyme is on top and the Fv of the antibody is at the bottom; the V_Ls are drawn with thinner lines on the left. The figures are oriented such that the Fvs are maximally superposed.

Fig. 10). There is complementarity in the antibody and antigen surfaces which are in contact, so that depressions in one are by and large filled by protrusions from the other [in D1.3 and HyHEL-5, solvent molecules are found trapped in the interface between the two macromolecules (Sheriff *et al.*, 1987; Fischmann *et al.*, 1991)]. There is also complementarity in the physical and chemical properties of the interacting surfaces, so that hydrogen bonds are formed whenever possible and, in the case of HyHEL-5, oppositely-charged side chains form ion pairs (Table 13b). Many aromatic residues figure prominently in these antibody–antigen interactions.

The interactions between antibody and specific ligand in various complexes are summarized in Table 14. The contacts involve van der Waals' interactions, hydrogen bonds between polar groups and ion pairs. Interestingly, the hydrogen bonds mostly involve side chain atoms; very few main-chain–main-chain hydrogen bonds are observed in these antibody–ligand complexes.

The interaction between antibody and larger ligands, e.g. peptides and trinucleotides, is comparable to that between antibody and whole antigen (Table 14). It is seen in Table 14, that antibody–peptide interactions approximate those between antibody and whole protein antigens, in terms of the surface areas buried upon complexation, the number of van der Waals' contacts and the number of hydrogen bonds formed. Details of the contact between B13I2 and a myohemerythrin-peptide homolog are included in Table 13 for comparison.

In the three anti-lysozymes, HyHEL-10, HyHEL-5 and D1.3, the surface that is used in the binding to antigen represents only 28.4, 26.5 and 21.4%, respectively, of the total surface formed by the CDRs; the CDR residues which contact the antigen represent only 35.8, 37.0 and 25.0% of the total number of CDR residues. The antibody–antigen contact utilizes primarily the central portion of the CDR surface in the three anti-lysozyme cases, and in all of the antibody–ligand structures determined so far.

The CDR residues involved in the contact with ligand in various antibody–ligand complexes of known structure are shown in Table 15. Many aromatic residues are seen to be involved in the contacts, whereas the involvement of small, apolar aliphatics is rare, a trend that was noted previously [Padlan (1990*a*); see also Mian *et al.* (1991)]. The ligand-contacting residues are seen to originate mostly from the C-terminal part of CDR1-L, the first and sometimes also the middle position in CDR2-L, from the whole of CDR3-L, CDR1-H and CDR3-H, and from the N-terminal part and the middle of CDR2-H. Thus, it is the central portion of the CDR surface (Fig. 4) that contacts the ligand in these, and possibly in all, antibody–antigen complexes.

The CDR3-L and CDR3-H are seen to play a prominent role, not only in ligand binding (Table 15), but also in the contact with the opposite domain (Table 8) and in the contact with the other CDRs (Table 6). It is probably no coincidence that these two CDR3s also usually display the greatest variability (Kabat *et al.*, 1991) and, in CDR3-H, the greatest variation in length and conformation (e.g. Wu *et al.*, 1993).

Although antigen binding primarily involves the CDRs, framework residues have been found on occasion to be involved also in the interaction with ligand. Thus the framework residue, Tyr49-L in D1.3 (Amit *et al.*, 1986; Fischmann *et al.*, 1991), and the Tyr49-L also in NC41 (Tulip *et al.*, 1992*a*), were found to be in contact with the antigen, as were the framework residues, Trp47-H in HyHEL-5 (Sheriff *et al.*, 1987) and Thr30-H in HyHEL-10 (Padlan *et al.*, 1989).

(ii) *The diversity of antigen-binding specificities.* The antigen-binding specificity of an antibody is defined by the physical and chemical properties of its CDR surface;

Table 14. Various antibody–ligand interactions

Antibody	Surf. buried	Ligand	Surf. buried	vdW	m.c.vdW	H.b.	m.c.H.b.	I.p.
McPC603 Fab	151	Phosphocholine	138	54	4	3		2
AN02 Fab	363	DNP-spin-label	230	129	5	2		0
4-4-20 Fab	338	Fluorescein	247	107	36	3		0
BV04-01 Fab	523	$d(pT)_3$	443	184	48	12		1
B13I2 Fab	503	Myohemerythrin peptide	439	113	27	13	2	2
17/9 Fab[a1]	488	Hemagglutinin peptide	418	143	36	14	1	1
Fab[a2]	536	Hemagglutinin peptide	459	154	34	13	1	1
Fab[b]	488	Hemagglutinin peptide	417	144	41	12	1	0
D1.3 Fab	537	Lysozyme	541	110	18	14	2	0
HyHEL-5 Fab	744	Lysozyme	741	145	25	13	0	3
HyHEL-10 Fab	716	Lysozyme	759	180	21	19	1	0
NC41 Fab	882	Neuraminidase	838	154	20	13	1	0

[a]From PDB Entry: 1HIM ([1]first Fab in the entry; [2]second Fab in the entry).
[b]From PDB Entry: 1HIN.
The contacts involving main chain atoms (m.c.) are listed separately. See footnote to Table 4 for the definition of contacts.

Table 15. CDR residues in contact with ligand in murine and human antigen-binding regions of known three-dimensional structure

	CDR1-L	CDR2-L	CDR3-L	CDR1-H	CDR2-H	CDR3-H
HyHEL-10	RASQ------SIGNNLH	YASQSIS	QQSNSWP-YT	SDYWS-	YVS---YSGSTYYNPSLKS	MDG------DY
HyHEL-5	SASS--------SVNYMY	DTSKLAS	QQWGRN--PT	DYWIE-	EILP--GSGSTNYHERFKG	GNYDF----DG
D1.3	RASG------NIHNYLA	YTTTLAD	QHFWSTP-RT	GYGVN-	MIW---GDGNTDYNSALKS	ERDYRL---DY
McPC603	KSSQSLLNSGNQKNFLA	GASTRES	QNDHSYP-LT	DFYME-	ASRNKGNKYTTEYSASVKG	NYYGSTWYFDV
4-4-20	RSSQSLVHS-QGNTYLR	KVSNRFS	SQSTHVP-WT	DYWMN-	QIRNKPYNYETYYSDSVKG	SYYGM----DY
AN02	SASS--------SVYYMY	DTSNLAS	QQWSSYPPIT	SDYAWN	YMS---YSGSTRYNPSLRS	GWPL-----AY
B1312[a]	RSNQTILLS-DGDTYLE	KVSNRFS	FQGSHVP-PT	RCAMS-	GISS--GGSYTFYPDTVKG	YSSDPFMF-DY
BV04-01	RSSQSLVHS-NGNTYLH	KVSNRFS	SQSTHVP-LT	TNAMN-	RIRSKSNNYATYYADSVKD	DQTGTAWF-AY
17/9[b]	TSSQSLFNSGKQKNYLT	WASTRES	QNDYSNP-LT	SYGMS-	TISN--GGGYTYYPDSVKG	RERYDENGFAY
MC41	KASQ------DVSTAVV	WASTRHI	QQHYSPP-WT	NYGMN-	WINT--NTGEPTYGEEFKG	GEDNFGSL-DY

[a] From PDB Entry: 2IGF.
[b] From PDB Entry: 1HIM (first Fab in the entry).
See footnote to Table 4 for the definition of contacts.

these in turn are determined by the conformation of the individual CDRs, by the relative disposition of the CDRs, and by the nature and disposition of the side chains of the amino acids in the CDRs. Thus, the pronounced structural variability in the CDRs, which is due not only to sequence variability, but also to the insertions and deletions frequently found in these segments, and the possible pairing of different V_L and V_H provide a ready explanation for the wide diversity of antigen-binding specificities. Within each CDR, there are residue positions that are more hypervariable than others; these positions are presumably more involved in the determination of antigen-binding specificity and in the diversification of specificities. The other residues in the CDRs may play an ancillary role and simply provide additional "stickiness" (see below). In addition, the more conserved residues in the CDRs may play a structural role and serve to stabilize the combining-site structure (Kabat *et al.*, 1977; Padlan, 1977*b*).

The possibility of "induced fit" provides an additional means of generating other specificities, but in this case without introducing any new ingredients into the picture. "Induced fit" can be achieved by small movements of side chains, or by more substantial structural modifications like the deformation of the CDR loops, or by a change in the relative disposition of the variable domains. The flexibility of the combining site, which allows the occurrence of "induced fit" necessarily results in the entropic loss upon complexation, but the greater interaction due to a more precise fit may result in an overall increase in binding energy.

Evidence is accumulating which suggests that any macromolecule, if presented properly, can be antigenic and, further, that all accessible areas of a macromolecule can potentially be bound by antibody [see Benjamin *et al.* (1984) for a review]. Moreover, there appear to be no special structural requirements for antigenicity (Padlan, 1992). If antibody combining sites are to participate in strong interactions with many possible antigenic structures, they must possess structural features which make them especially suited for interacting with ligands.

A survey (Padlan, 1990*a*) of the then available primary and three-dimensional data on the CDRs revealed that Tyr, His and Asn have a propensity for being in the CDRs. Moreover, although the exposure patterns of the various amino acid types in immunoglobulins are comparable to those in other water-soluble proteins, those with aromatic side chains are more exposed when in CDRs than when in the framework regions. Those results are confirmed when the now larger structural database on combining-site structures is analysed. The solvent accessibilities of the side chains in various Fv fragments of known three-dimensional structure are presented in Table 16a; the solvent exposures of the CDR and framework residues are presented separately in Tables 16b and 16c, respectively. For comparison, the solvent accessibilities of the side chains in 50 highly-refined structures of water-soluble proteins (Padlan, 1990*a*) are provided in Table 17.

Again, Trp and Tyr are found to be more buried when in the framework and more exposed when in the CDRs than in water-soluble proteins in general (Fig. 11); phenylalanines also seem to be following the trend. The Trp, Tyr and Phe in the framework regions of antibody variable domains are usually found in the interdomain interface (Fig. 5) and in the domain interiors [see also Poljak *et al.* (1975*a*), Davies *et al.* (1975*b*) and Novotny *et al.* (1983)]. When in the CDRs, these aromatic residues are frequently found to be involved in the interaction with ligand (see above and Fig. 10).

Amino acids with aromatic side chains can contribute significantly to ligand binding because of their large size (for greater hydrophobic effect), their large polarizabilities (for greater contribution to the van der Waals'

Table 16a. Exposure of residues in murine and human Fvs of known three-dimensional structure

	Bu	mB	pB	mE	Ex	Total
ALA	111	21	17	16	75	240
ARG	37	40	31	45	32	185
ASN	23	13	30	50	30	146
ASP	45	17	32	49	61	204
CYS	82	0	1	0	0	83
GLN	67	48	16	63	67	261
GLU	19	3	21	59	56	158
GLY	117	0	0	0	329	446
HIS	9	1	12	5	6	33
ILE	122	19	17	7	12	177
LEU	227	34	37	27	26	351
LYS	4	7	24	87	84	206
MET	59	5	8	5	0	77
PHE	118	14	12	7	3	154
PRO	25	10	30	76	43	184
SER	32	21	49	167	334	603
THR	46	36	82	132	91	387
TRP	85	10	9	5	0	109
TYR	114	54	55	35	15	273
VAL	162	53	21	16	28	280
Total	1504	406	504	851	1292	4557

interaction), and their ability to hydrogen bond [through their aromatic rings (Levitt and Perutz, 1988) or through polar atoms in their side chain], and, very importantly, because of their relative rigidity since they have few degrees of freedom (for lesser loss of conformational entropy upon complexation).

On the other hand, the apolar, aliphatic side chains of Val, Ile and Leu are incapable of participating in ionic interactions. They can contribute to ligand binding only through van der Waals' interactions and the hydrophobic effect. Moreover, in view of their rotational degrees of freedom which could be frozen upon complex formation, they can have a significant negative contribution to ligand binding (Novotny *et al.*, 1989; Padlan, 1990*a*).

It should also be noted that asparagines are more buried when in the CDRs than when in the framework. In the CDRs, the asparagines are found to be often involved in hydrogen bonds, many to main chain atoms. It was argued (Padlan, 1990*a*) that these asparagines serve to help stabilize the CDR loops (the main stabilizing factor would be the strong architecture of the domain framework) and permit the exposure of the hydrophobic aromatics, which otherwise would tend to be buried in the domain interior.

Thus the high incidence of exposed aromatic residues, as well as the paucity of apolar, aliphatic side chains, on part of its surface would give an antibody a "sticky patch" and give the molecule the capacity to bind diverse ligands. Specificity for a particular antigen would arise from the precise complementarity of the interacting surfaces and the correct positioning of complementary polar, esp. charged, groups on those surfaces (Levy *et al.*, 1989).

(iii) *Possible structural correlates of polyreactivity.* The affinity of polyreactive antibodies for the antigens to which they bind is rather low, with affinity constants ranging from 10^{-4} to 10^{-7} M (to be contrasted with monoreactive antibodies which show affinities for their specific ligands of 10^{-7} M or better) [see, for example, Burastero *et al.* (1988)]. Binding to many different ligands with low affinity may simply reflect the intrinsic stickiness of the antibody combining site. Indeed, it may be possible that an antibody that has been designated as being monospecific is actually capable of interacting with many different ligands, but that we are aware of only that one ligand for which the affinity is significantly higher than for the rest.

It should be pointed out that a 10-fold increase in affinity requires an improvement in the binding interaction of only 1.37 kcal/mole at room temperature which corresponds roughly to the energy of a hydrogen bond; a 1000-fold increase in affinity corresponds to an energy difference of 4.12 kcal/mole, or roughly the energy associated with a hydrogen bond that involves a charged group (Fersht *et al.*, 1985); the energies associated with salt bridges are difficult to estimate, but these would normally be higher and lead to greater increases in affinity than hydrogen bonds. This means that the difference in the reactivity patterns of polyreactive and monospecific antibodies may simply be a reflection of a single amino-acid difference.

Nevertheless, the concept of polyreactivity is intriguing and the existence of polyreactive natural antibodies clearly provides an organism the ability to cope, albeit weakly, with a deleterious antigen that it had not previously encountered (a better defense against subsequent antigenic challenge is then achieved by

Table 16b. Exposure of CDR residues in murine and human antigen-binding regions of known three-dimensional structure

	Bu	mB	pB	mE	Ex	Total
ALA	31	12	10	3	12	68
ARG	6	6	12	18	7	49
ASN	22	10	19	22	18	91
ASP	9	10	17	22	21	79
CYS	2	0	1	0	0	3
GLN	28	4	3	9	12	56
GLU	6	3	7	4	5	25
GLY	22	0	0	0	76	98
HIS	8	1	12	3	1	25
ILE	29	3	6	2	2	42
LEU	30	5	20	4	3	62
LYS	0	2	7	16	19	44
MET	16	1	0	2	0	19
PHE	16	8	8	5	1	38
PRO	7	10	5	14	6	42
SER	27	12	20	59	71	189
THR	11	15	31	20	4	81
TRP	6	8	9	5	0	28
TYR	13	36	39	35	15	138
VAL	24	9	2	3	4	42
Total	313	155	228	246	277	1219

Table 16c. Exposure of framework residues in murine and human antigen-binding regions of known three-dimensional structure

	Bu	mB	pB	mE	Ex	Total
ALA	80	9	7	13	63	172
ARG	31	34	19	27	25	136
ASN	1	3	11	28	12	55
ASP	36	7	15	27	40	125
CYS	80	0	0	0	0	80
GLN	39	44	13	54	55	205
GLU	13	0	14	55	51	133
GLY	95	0	0	0	253	348
HIS	1	0	0	2	5	8
ILE	93	16	11	5	10	135
LEU	197	29	17	23	23	289
LYS	4	5	17	71	65	162
MET	43	4	8	3	0	58
PHE	102	6	4	2	2	116
PRO	18	0	25	62	37	142
SER	5	9	29	108	263	414
THR	35	21	51	112	87	306
TRP	79	2	0	0	0	81
TYR	101	18	16	0	0	135
VAL	138	44	19	13	24	238
Total	1191	251	276	605	1015	3338

The fractional solvent accessibility values for the individual residues were computed as described by Padlan (1990*a*); residues, the sidechains of which have fractional accessibility values between 0.00 and 0.20, are designated as being completely buried (Bu), between 0.20 and 0.40 as mostly buried (mB), between 0.40 and 0.60 as partly buried/partly exposed (pB), between 0.60 and 0.80 as mostly exposed (mE), and at least 0.80 as completely exposed (Ex). In the special case of glycine, the residue is considered completely exposed if its α-carbon atom is accessible to solvent, otherwise it is considered completely buried. Here, residue exposures are defined in the context of an isolated Fv. The unliganded forms of McPC603, BV04-01, B13I2 (PDB Entry: 1IGF, first Fab in the entry) and 17/9 (PDB Entry: 1HIL, first Fab in the entry) were used in the computations.

affinity maturation). Several factors may contribute to polyreactivity and these are discussed below. Which, if any, can account for polyreactivity in antibodies may become obvious when sufficient structural data on these molecules become available.

It is possible, for example, that different areas of the antibody CDR surface are used for different antigens. In the case of the three anti-lysozymes examined above, less than a third of the CDR surface is utilized in the binding to the antigen so that many other antibody–ligand interactions can be envisioned. In addition, more of the framework residues could be recruited for involvement in the binding interaction.

A close fit between antibody and antigen is a feature of tight binding, and this may not be the case for polyreactivity. Departures from a close fit could lead to the presence of cavities in the interface, which is energetically expensive and will result in lower affinity. An imprecise fit between antibody combining site and antigenic determinant could be smoothed out by solvent molecules, but the immobilization of solvent molecules will result in a decrease in the entropy of the system and subtract from the total binding energy, again leading to lower affinity.

It may be that the combining site structure of a polyreactive antibody is unusually plastic so that the site could adopt many different conformations enabling it to accommodate many different antigenic structures. The plasticity of the combining site structure is mainly dependent on the deformability of the CDR loop structures and on the ability of V_L and V_H to assume different quaternary modes of association. The deformability of the CDRs will depend on their length and on the presence or absence of stabilizing interactions with neighboring structures, as well as on the presence or absence in these regions of certain amino acid types like glycine (which usually confers flexibility), proline (which usually confers rigidity), or asparagine (which, by hydrogen bonding to main chain atoms, contributes to the stability of the local structure). Greater plasticity, therefore, may be reflected in a more frequent occurrence of glycines and/or a reduced presence of prolines or asparagines. In this regard, it is interesting that, of the three possible reading frames, the one used to transcribe the D-gene segment is usually that which results in the presence of glycines in the CDR3-H (Abergel and Claverie, 1991).

It is also possible that a polyreactive combining site is unusually sticky, as a consequence of the presence or absence of certain amino acid types which give the site a greater capacity for binding ligands. Greater stickiness

Table 17. Exposure of amino acid residues in 50 highiy-refined water-soluble protein structures

	Bu	mB	pB	mE	Ex	Total
ALA	309	79	97	117	199	801
ARG	21	42	83	80	50	276
ASN	49	52	99	123	145	468
ASP	69	53	79	113	151	465
CYS	174	46	26	12	1	259
GLN	39	47	61	106	91	344
GLU	30	33	72	132	127	394
GLY	291	0	0	0	635	926
HIS	63	39	41	23	25	191
ILE	273	85	62	22	8	450
LEU	372	121	70	43	22	628
LYS	10	29	91	194	187	511
MET	88	20	17	13	6	144
PHE	177	70	42	8	8	305
PRO	81	32	54	86	109	362
SER	156	74	79	145	293	747
THR	140	68	123	145	135	611
TRP	68	44	14	7	3	136
TYR	106	113	74	48	20	361
VAL	385	101	84	61	36	667
Total	2915	1157	1268	1478	2251	9069

Here, a structure is considered to be a highly-refined structure if it has been determined to a resolution of 1.8 Å or better and refined to a crystallographic R-value of 0.200 or better. See footnote to Table 16.

Best Available Copy

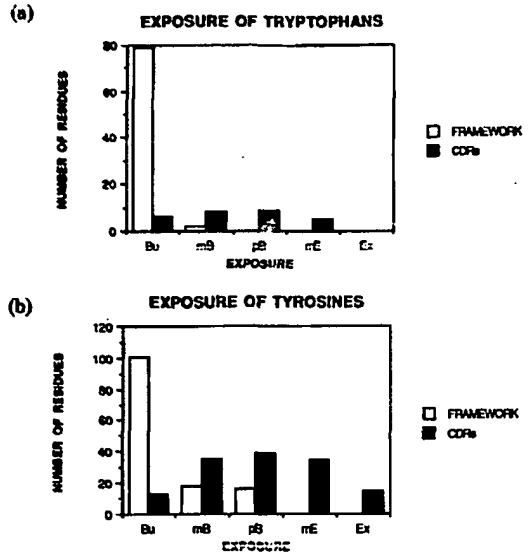


Fig. 11. Histograms showing the exposure patterns of the Trp (a) and Tyr (b) residues in antigen-binding regions of known three-dimensional structure. Shown are the numbers of these residues which are completely buried (Bu), mostly buried (mB), partly buried (pB), mostly exposed (mE) and completely exposed (Ex), as enumerated in Table 16.

may be reflected in an increased presence of aromatics vs aliphatic residues, for example. The observation that some polyreactive antibodies have many charged residues in and around their CDRs (Gonzalez-Quintial *et al.*, 1990) is intriguing.

Fc–ligand interactions

A number of molecules are known to bind to Fc, including low- and high-affinity Fc receptors, C1q and Protein A. For these, the location of their binding sites on the Fc has been deduced by sequence correlations, chemical modifications and site-directed mutagenesis, and, in the case of Protein A, by direct crystallographic analysis.

Duncan *et al.* (1988), by site-directed mutagenesis, have shown that Leu235 (Eu numbering) is a major determinant in the binding of murine Fc to the high-affinity receptor on monocytes. Unfortunately, the corresponding residue is part of the stretch that is disordered in the crystal structure of human IgG1 Fc and, therefore, cannot be located in that structure. The recent crystal structure analysis of the intact murine IgG2a antibody (Harris *et al.*, 1992), in which the full hinge is visible, should help in delineating the receptor-binding site on IgG.

It is known that C1q binds to the C_H2 domain and numerous attempts have been made to localize the site of binding [reviewed by Burton (1985)]. Recently, by systematically altering residues that are expected to lie on the surface, Duncan and Winter (1988) have been able to propose the location of the minimal residues that are probably involved in the binding of C1q; these are residues 318, 320 and 322 (Eu numbering). In human IgG1, these residues lie on the outside face of $C\gamma2$ (Fig. 12) and their disposition will allow a direct contact with C1q.

The crystallographic analysis of the binding of Fragment B of Protein A to human IgG1 Fc (Deisenhofer, 1981) demonstrated that the Protein A binding site is at the junction of the $C\gamma2$ and $C\gamma3$ domains (Fig. 12). Both domains are involved in the contact and approximately 1200 Å of accessible surface area on the Fc and the ligand are buried by the interaction. The contact is mainly hydrophobic with a few hydrogen bonds.

STRUCTURAL ASPECTS OF ANTIBODY APPLICATIONS

Antibodies of predefined specificity have many potential uses in industry and in medicine [see, for example, Steinman (1990), Schlom (1991), Waldmann (1991) and Co and Queen (1991)] and the advent of hybridoma technology (Koehler and Milstein, 1975) has made possible the generation of virtually limitless amounts of such antibodies. Monoclonal antibodies are being used for

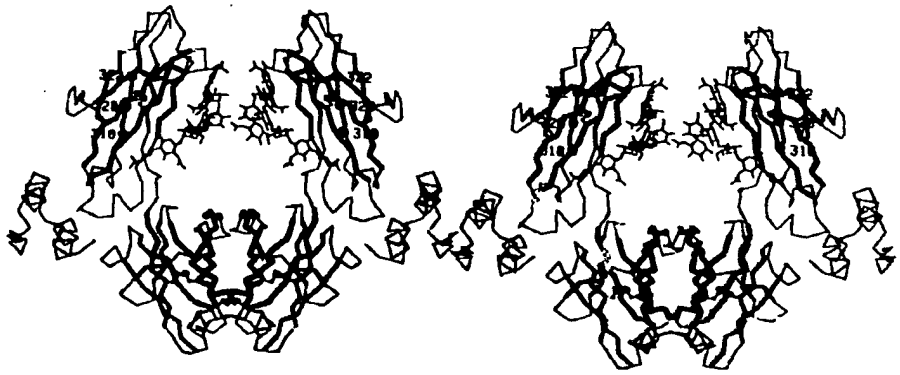


Fig. 12. Stereodrawing of the α-carbon trace of human IgG1 Fc showing various binding sites. The orientation is as in Fig. 7. Fragment B of Protein A is shown as the tri-helical structure near the junction of the C_H2 and C_H3 domains. Residues 318, 320 and 322, which have been identified as being involved in C1q binding (Duncan and Winter, 1988), are indicated by large, filled circles on C_H2.